Knowledge-Augmented Methods for Natural Language Processing

Chenguang Zhu[1], Yichong Xu[1], Xiang Ren[2], Bill Yuchen Lin[2], Meng Jiang[3], Wenhao Yu[3]
[1]Microsoft Cognitive Services Research, [2]University of Southern California,
[3]University of Notre Dame
{chezhu, yicxu}@microsoft.com, {xiangren, yuchen.lin}@usc.edu,
{mjiang2, wyu1}@nd.edu

1 Information

Keywords Knowledge, common sense, language understanding, language generation.

Tutorial description Knowledge in NLP has been a rising trend especially after the advent of large-scale pre-trained models. Knowledge is critical to equip statistics-based models with common sense, logic and other external information. In this tutorial, we will introduce recent state-of-the-art works in applying knowledge in language understanding, language generation and commonsense reasoning.

Suggested duration Half day (3 hours)

Type of Tutorial Cutting-edge

Targeted Audience Target audience are researchers and practitioners in natural language processing, knowledge graph and common sense reasoning. The audience will learn about the state-of-the-art research in integrating knowledge into NLP to improve the cognition capability of models.

Prerequisites Basic knowledge in machine learning and natural language processing. Optional background in knowledge graph.

Outline

- Introduction to NLP and Knowledge (15 min)
- Knowledge in Natural Language Understanding (55 min)
- Knowledge in Natural Language Generation (55 min)
- Commonsense Knowledge and Reasoning for NLP (55 min)

Estimated number of participants 150.

Preferable venues ACL / NAACL-HLT / EMNLP

Similar tutorials There have been several tutorials/workshops on knowledge in NLP:

- Tutorial at AAAI 2021: Commonsense Knowledge Acquisition and Representation
- Tutorial at EMNLP 2021: Knowledge-Enriched Natural Language Generation
- KR2ML workshop at NeurIPS 2019 and 2020: Knowledge Representation & Reasoning Meets Machine Learning
- Tutorial at ACL 2020: Commonsense Reasoning for Natural Language Processing

Diversity considerations The use of knowledge is not limited to any specific language. The technologies we introduce are generally applicable to all languages, as long as there is corresponding corpus and knowledge sources, e.g., dictionaries, knowledge graph, etc. We have a diverse instructor team across multiple institutions (i.e., MS, USC, UND). The team has a diverse and broad expertise in natural language processing and generation, machine learning, and various application domains.

Ethics Existing corpus and knowledge graphs may contain varying degrees of bias. Thus, the usage of related technologies in applications should come with modules to detect and correct such biases, especially for language generation applications.

2 Brief Tutorial Outline

In recent years, the field of natural language processing has considerably benefited from larger-scale models, better training strategies, and greater availability of data, exemplified by BERT* (Devlin et al., 2019), RoBERTa* (Liu et al., 2019b), and GPT models (Radford et al., 2018, 2019; Brown et al., 2020). It has been shown that these pre-trained language models can effectively characterize linguistic patterns in text and generate high-quality context-aware representations (Liu et al., 2019a). However, these models are trained in a way where the only input is the source text. As a result, these models struggle to grasp external world knowledge about concepts, relations, and common sense (Poerner et al., 2019; Talmor et al., 2020).

In this tutorial, we use *Knowledge* to refer to this external information which is absent from model input yet useful for the model to produce target output. Knowledge is important for language representation and should be included into the training and inference of language models. Knowledge is also an indispensable component to enable higher levels of intelligence which is unattainable from statistical learning on input text patterns.

2.1 Knowledge-augmented Natural Language Understanding

In natural language understanding (NLU), the task is to make predictions about the property of words, phrases, sentences or paragraphs based on the input text, e.g., sentiment analysis, named entity recognition and language inference. We will introduce how to use knowledge to augment NLU models along the dimension of knowledge source: i) structured knowledge such as knowledge graph, and ii) unstructured knowledge such as text corpus.

We first discuss efforts to integrate structured knowledge into language understanding, which can be categorized into explicit methods via concept/entity embeddings (Zhang et al., 2019; Peters et al., 2019; Liu et al., 2020; Yu et al., 2020a) and implicit methods via entity masking prediction (Sun et al., 2019; Shen et al., 2020; Xiong et al., 2020; Wang et al., 2019). For example, ERNIE* (Zhang et al., 2019) explicitly pre-trains the entity embeddings on a knowledge graph using TransE (Bordes et al., 2013), while EAE (Févry et al., 2020) learns the representation as model parameters. KEPLER (Wang et al., 2019) implicitly calculates entity embeddings using a pre-trained language model based on the description text. Recently, some works propose to co-train the knowledge graph module and the language model (Ding et al., 2019; Lv et al., 2020; Yu et al., 2022b). For example, JAKET* (Yu et al., 2022b) proposes to use the knowledge module to produce embeddings for entities in text while using the language module to generate context-aware initial embeddings for entities and relations in the knowledge graph. Yu et al. (2021a) and Xu et al. (2021)* propose to use dictionary descriptions as additional knowledge source for natural language understanding and commonsense reasoning tasks.

We then introduce how to integrate unstructured knowledge into NLU models. This usually requires a text retrieval module to obtain related text from knowledge corpus. There have been multiple approaches to adopt unstructured knowledge, especially for open-domain QA task. For example, Lee et al. (2019) first trains a retriever by inverse cloze task (ICT) and then jointly trains the retriever and reader for open-domain QA. DPR* (Karpukhin et al., 2020) conducts supervised training for the retriever and achieves better performance on open-domain QA. REALM (Guu et al., 2020) predicts masked salient spans consisting of entities to jointly pre-train the reader and retriever. KG-FiD (Yu et al., 2022a) proposed to filter noisy passages by leveraging the structural relationship among the retrieved passages with a knowledge graph during retrieval.

We will introduce the above methods and focus on three key aspects of employing knowledge in NLU tasks: i) how to ground the input into knowledge domain (e.g., entity linking), ii) how to represent knowledge (e.g., graph neural network), and iii) how to integrate knowledge information into the NLU models (e.g., attention).

2.2 Knowledge-augmented Natural Language Generation

The goal of natural language generation (NLG) is to produce understandable text in human language from linguistic or non-linguistic data in a variety of forms such as textual data, image data, and structured knowledge graph (Yu et al., 2020b). Different from natural language understanding (NLU) methods, NLG methods are typically under the encoder-decoder generation framework (Sutskever et al., 2014; Bahdanau et al., 2015), which poses unique challenges for leveraging knowledge into decoding the next tokens during generation.

We will first present the existing methods for integrating knowledge into NLG models. These models are categorized into three major paradigms which incorporate knowledge through (1) *model architectures* that facilitate the use of knowledge, such as knowledge-related attention mechanism, knowledge-related copy/pointer mechanisms (Zhou et al., 2018; Zhang et al., 2020a; Liu et al., 2021a; Guan et al., 2020a; Dong et al., 2021); (2) *learning frameworks* that inject knowledge information into the generation models through training, such as posterior regularization, constraint-driven learning, semantic loss, knowledge-informed weak supervision (Hu et al., 2016, 2018; Tan et al., 2020; Dinan et al., 2019); (3) *inference methods* which imposes on the inference process different knowledge constraints to guide decoding, such as lexical constraints, task-specific objectives, global inter-dependency (Dathathri et al., 2020; Qin et al., 2020; Yu et al., 2021b).

In addition to presenting the unified model architectures/frameworks, we will introduce several specific methods based on different knowledge sources. The knowledge sources can be divided into structured knowledge such as knowledge graph, or unstructured such as text corpus. Many methods have been proposed to learn the relationship between structured knowledge and input/output sequences. They can be categorized into four methodologies: injecting pre-computed knowledge embeddings into language generation (Zhou et al., 2018); transferring knowledge into language model with triplet information (Guan et al., 2020a); performing reasoning over knowledge graph via path finding strategies (Liu et al., 2019c; Ji et al., 2020a); and improve the graph embeddings with graph neural networks (Zhang et al., 2020a; Ji et al., 2020b). For example, Zhou et al. (2018) enriched the context representations of the input sequence with neighbouring concepts on ConceptNet using graph attention. Recently, some work attempted to integrate external commonsense knowledge into generative pretrained language models (Guan et al., 2020a; Bhagavatula et al., 2020). For example, Guan et al. (2020a) conducted post-training on synthetic data constructed from commonsense KG by translating triplets into natural language texts.

To handle different kinds of relationships between unstructured text and input/output sequences, existing methods can be categorized into two methodologies: guiding generation with retrieved information (Ghazvininejad et al., 2018; Lewis et al., 2020; Wang et al., 2021); modeling background knowledge into text generation (Qin et al., 2019; Meng et al., 2020; Zeng et al., 2021). For example, Lewis et al. (2020) introduced a general retrieval-augmented generation (RAG) framework by leveraging a pre-trained neural retriever and generator. It can be easily fine-tuned on downstream tasks, and it has demonstrated state-of-the-art performance on various knowledge-intensive natural language generation tasks.

2.3 Commonsense Knowledge and Reasoning for Natural Language Processing

Humans reason and make decisions in everyday settings by using *common sense*, which consists of basic knowledge (*e.g.*, regarding the physical world or human social behavior) that is rarely taught explicitly yet shared by almost everyone. Commonsense knowledge and the ability of using common sense to reason is thus of vital significance for developing human-like NLP models as well as general-purpose AI systems. We will cover topics as follows: (1) **resources** and **datasets** for developing and benchmarking commonsense reasoning methods. (2) **knowledge-aware commonsense reasoning methods** for both understanding and generation tasks. (3) **analysis** on the acquired commonsense knowledge of pre-trained LMs and the behavior of knowledge-augmented commonsense reasoning methods.

There is a recent surge of novel knowledge resources and the benchmark datasets for researching commonsense in the NLP domain. One of the most widely used commonsense knowledge resource is ConceptNet (Speer et al., 2017), which is a binary, relational knowledge graph. Although ConceptNet enjoys simplicity and popularity, its incompleteness and concept-centric structures limit the development of more general topics on commonsense reasoning for NLP. We present the recent works on developing commonsense knowledge resources, such as ASER (Zhang et al., 2021), AscentKB (Nguyen et al., 2021), COMET-ATOMIC2020 (Hwang et al., 2021), and GenericsKB (Bhakthavatsalam et al., 2020), which provide us with event-centric, large-scale, neural-symbolic, semi-structured ways to access and model commonsense knowledge. We then introduce the popular datasets for evaluating the commonsense reasoning methods that span three main categories: 1) multiple-choice QA (e.g., CommonsenseQA (Talmor et al., 2019), SocialIQA (Sap et al., 2019), PhysicalIQA (Bisk et al., 2020), RiddleSense (Lin et al., 2021b)), 2) open-ended QA (e.g., ProtoQA (Boratko et al., 2020) OpenCSR (Lin et al., 2021a)), 3) constrained NLG (e.g., CommonGen (Lin et al., 2020b), conversation generation, and story generation).

To equip language models (LMs) with commonsense reasoning ability, researchers have developed many knowledge-augmented reasoning models that fit different task formulations. For the multiple-choice QA setting, we introduce a set of knowledge-augmented neuro-symbolic methods: KagNet* (Lin et al., 2019), HyKAS (Ma et al., 2019), MHGRN* (Feng et al., 2020), HybridGN (Yan et al., 2020) and QA-GNN* (Yasunaga et al., 2021). These methods make use of structured knowledge graphs and/or neural commonsense KBs for injecting external knowledge structures to neural LMs. As for the open-ended setting, we present the DrKIT (Dhingra et al., 2020) and DrFact* (Lin et al., 2021a) reasoning frameworks, which are both designed for differentiable reasoning over a virtual knowledge graph (i.e., an un/semi-structured text corpus).

For generation-based commonsense tasks, we present knowledge-augmented text generation models that are designed for generative commonsense: 1) EKI-BART (Fan et al., 2020), KG-BART* (Liu et al., 2021b), and RE-T5* (Wang et al., 2021) for the CommonGen task, 2) commonsense knowledge-enhanced story generation models (Guan et al., 2019, 2020b), and 3) commonsense-based models for conversation generation, such as ConceptFlow* (Zhang et al., 2020b) and CARE (Zhong et al., 2021).

Apart from the benchmarking and modeling, we also introduce the analysis works that aim to provide a deeper understanding the commonsense knowledge of pre-trained LMs: LAMA Probing* (Petroni et al., 2019), NumerSense (Lin et al., 2020a), and RICA* (Zhou et al., 2020). In addition, we also introduce the line of works that focus on interpreting the reasoning mechanism of the knowledge-augmented reasoning methods (Raman et al., 2021; Chan et al., 2021; Rajani et al., 2019).

3 Presenters

Chenguang Zhu is a Principal Research Manager in Microsoft Cognitive Services Research Group, where he leads the Knowledge & Language Team. His research in NLP covers knowledge graph, text summarization and task-oriented dialogue. Dr. Zhu has led teams to achieve first places in multiple NLP competitions, including CommonsenseQA, CommonGen, FEVER, CoQA, ARC and SQuAD v1.0. He holds a Ph.D. degree in Computer Science from Stanford University. Dr. Zhu has given talks at Stanford University, Carnegie Mellon University and University of Notre Dame. He has previously been TA for Coursera online class "Automata", giving teaching sessions to 100K international students. Additional information is available at https://www.microsoft.com/en-us/research/people/chezhu/.

Yichong Xu is a Senior Researcher in Knowledge & Language Team in Microsoft Cognitive Services Research Group. His research in NLP focuses on using external knowledge to help natural language processing, including question answering, commonsense reasoning, and text summarization. Dr. Xu received his Ph.D. in Machine Learning from Carnegie Mellon University. During his time at CMU, he has been TA for large classes ($>$ 200 students) on machine learning and convex optimization. Dr. Xu has given talks at CMU AI Seminar, as well as in many international conferences including ACL, NAACL, NeurIPS, ICML, etc. Additional information is available at https://xycking.wixsite.com/yichongxu.

Xiang Ren is an assistant professor at the USC Computer Science Department, a Research Team Leader at USC ISI, and the PI of the Intelligence and Knowledge Discovery (INK) Lab at USC. Priorly, he received his Ph.D. in Computer Science from the University of Illinois Urbana-Champaign. Dr. Ren works on knowledge acquisition and reasoning in natural language processing, with focuses on developing human-centered and label-efficient computational methods for building trustworthy NLP systems. Ren publishes over 100 research papers and delivered over 10 tutorials at the top conferences in natural language process, data mining, and artificial intelligence. He received NSF CAREER Award, The Web Conference Best Paper runner-up, ACM SIGKDD Doctoral Dissertation Award, and several research awards from Google, Amazon, JP Morgan, Sony, and Snapchat. He was named Forbes' Asia 30 Under 30 in 2019. Additional information is available at https://shanzhenren.github.io/.

Bill Yuchen Lin is a Ph.D. candidate at USC. His research goal is to teach machines to think, talk, and act with commonsense knowledge and commonsense reasoning ability as humans do. Towards this ultimate goal, he has been developing knowledge-augmented reasoning methods (e.g., KagNet, MHGRN, DrFact) and constructing benchmark datasets (e.g., CommonGen, RiddleSense, X-CSR) that require commonsense knowledge and complex reasoning for both NLU and NLG. He initiated an online compendium of commonsense reasoning research, which serves as a portal site[1] for the community. More information is available at https://yuchenlin.xyz/.

Meng Jiang is an assistant professor in the Department of Computer Science and Engineering at the University of Notre Dame. He received his B.E. and Ph.D. in Computer Science from Tsinghua University and was a postdoctoral research associate at the University of Illinois at Urbana-Champaign. His research interests focus on knowledge graph construction and natural language generation for news summarization and forum post generation. The awards he received include Notre Dame Faculty Award in 2019 and Best Paper Awards at ISDSA and KDD-DLG in 2020. Additional information is available at http://www.meng-jiang.com/.

Wenhao Yu is a Ph.D. student in the Department of Computer Science and Engineering at the University of Notre Dame. His research lies in controllable knowledge-driven natural language processing, particularly in natural language generation. His research has been published in top-ranked NLP and data mining conferences such as ACL, EMNLP, KDD and WWW. Additional information is available at https://wyu97.github.io/.

[1] https://commonsense.run/

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Chandra Bhagavatula, Ronan Le Bras, Chaitanya Malaviya, Keisuke Sakaguchi, Ari Holtzman, Hannah Rashkin, Doug Downey, Wen-tau Yih, and Yejin Choi. 2020. Abductive commonsense reasoning. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Sumithra Bhakthavatsalam, Chloe Anastasiades, and P. Clark. 2020. Genericskb: A knowledge base of generic statements. *ArXiv*, abs/2005.00660.

Yonatan Bisk, Rowan Zellers, Ronan LeBras, Jianfeng Gao, and Yejin Choi. 2020. PIQA: reasoning about physical commonsense in natural language. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 7432–7439. AAAI Press.

Michael Boratko, Xiang Li, Tim O'Gorman, Rajarshi Das, Dan Le, and Andrew McCallum. 2020. ProtoQA: A question answering dataset for prototypical common-sense reasoning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1122–1136, Online. Association for Computational Linguistics.

Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2787–2795.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Aaron Chan, Soumya Sanyal, Bo Long, Jiashu Xu, Tanishq Gupta, and Xiang Ren. 2021. Salkg: Learning from knowledge graph explanations for commonsense reasoning. *ArXiv*, abs/2104.08793.

Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. 2020. Plug and play language models: A simple approach to controlled text generation. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Bhuwan Dhingra, Manzil Zaheer, Vidhisha Balachandran, Graham Neubig, Ruslan Salakhutdinov, and William W. Cohen. 2020. Differentiable reasoning over a virtual knowledge base. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, and Jason Weston. 2019. Wizard of wikipedia: Knowledge-powered conversational agents. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

Ming Ding, Chang Zhou, Qibin Chen, Hongxia Yang, and Jie Tang. 2019. Cognitive graph for multi-hop reading comprehension at scale. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2694–2703, Florence, Italy. Association for Computational Linguistics.

Xiangyu Dong, Wenhao Yu, Chenguang Zhu, and Meng Jiang. 2021. Injecting entity types into entity-guided text generation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Zhihao Fan, Yeyun Gong, Zhongyu Wei, Siyuan Wang, Yameng Huang, Jian Jiao, Xuanjing Huang, Nan Duan, and Ruofei Zhang. 2020. An enhanced knowledge injection model for commonsense generation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2014–2025, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Yanlin Feng, Xinyue Chen, Bill Yuchen Lin, Peifeng Wang, Jun Yan, and Xiang Ren. 2020. Scalable multi-hop relational reasoning for knowledge-aware question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1295–1309, Online. Association for Computational Linguistics.

Thibault Févry, Livio Baldini Soares, Nicholas FitzGerald, Eunsol Choi, and Tom Kwiatkowski. 2020. Entities as experts: Sparse memory access with entity supervision. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4937–4951, Online. Association for Computational Linguistics.

Marjan Ghazvininejad, Chris Brockett, Ming-Wei Chang, Bill Dolan, Jianfeng Gao, Wen-tau Yih, and Michel Galley. 2018. A knowledge-grounded neural conversation model. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 5110–5117. AAAI Press.

Jian Guan, Fei Huang, Zhihao Zhao, Xiaoyan Zhu, and Minlie Huang. 2020a. A knowledge-enhanced pretraining model for commonsense story generation. *Transactions of the Association for Computational Linguistics*, 8:93–108.

Jian Guan, Fei Huang, Zhihao Zhao, Xiaoyan Zhu, and Minlie Huang. 2020b. A knowledge-enhanced pretraining model for commonsense story generation. *Transactions of the Association for Computational Linguistics*, 8:93–108.

Jian Guan, Yansen Wang, and Minlie Huang. 2019. Story ending generation with incremental encoding and commonsense knowledge. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 6473–6480. AAAI Press.

Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. 2020. Realm: Retrieval-augmented language model pre-training. *arXiv preprint arXiv:2002.08909*.

Zhiting Hu, Xuezhe Ma, Zhengzhong Liu, Eduard Hovy, and Eric Xing. 2016. Harnessing deep neural networks with logic rules. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2410–2420, Berlin, Germany. Association for Computational Linguistics.

Zhiting Hu, Zichao Yang, Ruslan Salakhutdinov, Lianhui Qin, Xiaodan Liang, Haoye Dong, and Eric P. Xing. 2018. Deep generative models with learnable knowledge constraints. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 10522–10533.

Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, and Yejin Choi. 2021. Comet-atomic 2020: On symbolic and neural commonsense knowledge graphs. In *AAAI*.

Haozhe Ji, Pei Ke, Shaohan Huang, Furu Wei, and Minlie Huang. 2020a. Generating commonsense explanation by extracting bridge concepts from reasoning paths. In *Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and International Joint Conference on Natural Language (AACL-IJCNLP)*.

Haozhe Ji, Pei Ke, Shaohan Huang, Furu Wei, Xiaoyan Zhu, and Minlie Huang. 2020b. Language generation with multi-hop reasoning on commonsense knowledge graph. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 725–736, Online. Association for Computational Linguistics.

Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6769–6781, Online. Association for Computational Linguistics.

Kenton Lee, Ming-Wei Chang, and Kristina Toutanova. 2019. Latent retrieval for weakly supervised open domain question answering. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6086–6096, Florence, Italy. Association for Computational Linguistics.

Patrick S. H. Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Bill Yuchen Lin, Xinyue Chen, Jamin Chen, and Xiang Ren. 2019. KagNet: Knowledge-aware graph networks for commonsense reasoning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2829–2839, Hong Kong, China. Association for Computational Linguistics.

Bill Yuchen Lin, Seyeon Lee, Rahul Khanna, and Xiang Ren. 2020a. Birds have four legs?! NumerSense: Probing Numerical Commonsense Knowledge of Pre-Trained Language Models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6862–6868, Online. Association for Computational Linguistics.

Bill Yuchen Lin, Haitian Sun, Bhuwan Dhingra, Manzil Zaheer, Xiang Ren, and William Cohen. 2021a. Differentiable open-ended commonsense reasoning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4611–4625, Online. Association for Computational Linguistics.

Bill Yuchen Lin, Ziyi Wu, Yichi Yang, Dong-Ho Lee, and Xiang Ren. 2021b. Riddlesense: Reasoning about riddle questions featuring linguistic creativity and commonsense knowledge. In *ACL*.

Bill Yuchen Lin, Wangchunshu Zhou, Ming Shen, Pei Zhou, Chandra Bhagavatula, Yejin Choi, and Xiang Ren. 2020b. CommonGen: A constrained text generation challenge for generative commonsense reasoning. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1823–1840, Online. Association for Computational Linguistics.

Nelson F. Liu, Matt Gardner, Yonatan Belinkov, Matthew E. Peters, and Noah A. Smith. 2019a. Linguistic knowledge and transferability of contextual representations. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1073–1094, Minneapolis, Minnesota. Association for Computational Linguistics.

Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. 2020. K-BERT: enabling language representation with knowledge graph. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 2901–2908. AAAI Press.

Ye Liu, Yao Wan, Lifang He, Hao Peng, and Philip S Yu. 2021a. Kg-bart: Knowledge graph-augmented bart for generative commonsense reasoning. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Ye Liu, Yao Wan, Lifang He, Hao Peng, and Philip S. Yu. 2021b. Kg-bart: Knowledge graph-augmented bart for generative commonsense reasoning. In *AAAI*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019b. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Zhibin Liu, Zheng-Yu Niu, Hua Wu, and Haifeng Wang. 2019c. Knowledge aware conversation generation with reasoning on augmented graph. In *Conference on Empirical Methods in Natural Language Processing and International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.

Shangwen Lv, Daya Guo, Jingjing Xu, Duyu Tang, Nan Duan, Ming Gong, Linjun Shou, Daxin Jiang, Guihong Cao, and Songlin Hu. 2020. Graph-based reasoning over heterogeneous external knowledge for commonsense question answering. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 8449–8456. AAAI Press.

Kaixin Ma, Jonathan Francis, Quanyang Lu, Eric Nyberg, and Alessandro Oltramari. 2019. Towards generalizable neuro-symbolic systems for commonsense question answering. In *Proceedings of the First Workshop on Commonsense Inference in Natural Language Processing*, pages 22–32, Hong Kong, China. Association for Computational Linguistics.

Chuan Meng, Pengjie Ren, Zhumin Chen, Christof Monz, Jun Ma, and Maarten de Rijke. 2020. Refnet: A reference-aware network for background based conversation. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Tuan-Phong Nguyen, Simon Razniewski, and G. Weikum. 2021. Advanced semantics for commonsense knowledge extraction. *Proceedings of the Web Conference 2021*.

Matthew E. Peters, Mark Neumann, Robert Logan, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A. Smith. 2019. Knowledge enhanced contextual word representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 43–54, Hong Kong, China. Association for Computational Linguistics.

Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. 2019. Language models as knowledge bases? In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2463–2473, Hong Kong, China. Association for Computational Linguistics.

Nina Poerner, Ulli Waltinger, and Hinrich Schütze. 2019. Bert is not a knowledge base (yet): Factual knowledge vs. name-based reasoning in unsupervised qa. *arXiv preprint arXiv:1911.03681*.

Lianhui Qin, Michel Galley, Chris Brockett, Xiaodong Liu, Xiang Gao, Bill Dolan, Yejin Choi, and Jianfeng Gao. 2019. Conversing by reading: Contentful neural conversation with on-demand machine reading. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Lianhui Qin, Vered Shwartz, Peter West, Chandra Bhagavatula, Jena D Hwang, Ronan Le Bras, Antoine

Bosselut, and Yejin Choi. 2020. Backpropagation-based decoding for unsupervised counterfactual and abductive reasoning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 794–805.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8):9.

Nazneen Fatema Rajani, Bryan McCann, Caiming Xiong, and Richard Socher. 2019. Explain yourself! leveraging language models for commonsense reasoning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4932–4942, Florence, Italy. Association for Computational Linguistics.

Mrigank Raman, Siddhant Agarwal, Peifeng Wang, Aaron Chan, Hansen Wang, Sungchul Kim, Ryan A. Rossi, Handong Zhao, Nedim Lipka, and Xiang Ren. 2021. Learning to deceive knowledge graph augmented models via targeted perturbation. *ArXiv*, abs/2010.12872.

Maarten Sap, Hannah Rashkin, Derek Chen, Ronan Le Bras, and Yejin Choi. 2019. Social IQa: Commonsense reasoning about social interactions. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4463–4473, Hong Kong, China. Association for Computational Linguistics.

Tao Shen, Yi Mao, Pengcheng He, Guodong Long, Adam Trischler, and Weizhu Chen. 2020. Exploiting structured knowledge in text via graph-guided representation learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8980–8994, Online. Association for Computational Linguistics.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, pages 4444–4451. AAAI Press.

Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. 2019. Ernie: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1904.09223*.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 3104–3112.

Alon Talmor, Yanai Elazar, Yoav Goldberg, and Jonathan Berant. 2020. oLMpics-on what language model pre-training captures. *Transactions of the Association for Computational Linguistics*, 8:743–758.

Alon Talmor, Jonathan Herzig, Nicholas Lourie, and Jonathan Berant. 2019. CommonsenseQA: A question answering challenge targeting commonsense knowledge. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4149–4158, Minneapolis, Minnesota. Association for Computational Linguistics.

Bowen Tan, Lianhui Qin, Eric Xing, and Zhiting Hu. 2020. Summarizing text on any aspects: A knowledge-informed weakly-supervised approach. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6301–6309, Online. Association for Computational Linguistics.

Han Wang, Yang Liu, Chenguang Zhu, Linjun Shou, Ming Gong Gong, Yichong Xu, and Michael Zeng. 2021. Retrieval enhanced model for commonsense generation. In *Annual Meeting of Association for Computational Linguistics (ACL)*.

Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhiyuan Liu, Juanzi Li, and Jian Tang. 2019. Kepler: A unified model for knowledge embedding and pre-trained language representation. *arXiv preprint arXiv:1911.06136*.

Wenhan Xiong, Jingfei Du, William Yang Wang, and Veselin Stoyanov. 2020. Pretrained encyclopedia: Weakly supervised knowledge-pretrained language model. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Yichong Xu, Chenguang Zhu, Ruochen Xu, Yang Liu, Michael Zeng, and Xuedong Huang. 2021. Fusing context into knowledge graph for commonsense reasoning. In *ACL*.

Jun Yan, Mrigank Raman, Tianyu Zhang, Ryan A. Rossi, Handong Zhao, Sungchul Kim, Nedim Lipka, and Xiang Ren. 2020. Learning contextualized knowledge structures for commonsense reasoning. *ArXiv*, abs/2010.12873.

Michihiro Yasunaga, Hongyu Ren, Antoine Bosselut, Percy Liang, and Jure Leskovec. 2021. QA-GNN: Reasoning with language models and knowledge graphs for question answering. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 535–546, Online. Association for Computational Linguistics.

Donghan Yu, Chenguang Zhu, Yuwei Fang, Wenhao Yu, Shuohang Wang, Yichong Xu, Xiang Ren, Yiming Yang, and Michael Zeng. 2022a. Kg-fid: Infusing knowledge graph in fusion-in-decoder for open-domain question answering. *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Donghan Yu, Chenguang Zhu, Yiming Yang, and Michael Zeng. 2022b. Jaket: Joint pre-training of knowledge graph and language understanding. *AAAI Conference on Artificial Intelligence (AAAI)*.

Wenhao Yu, Mengxia Yu, Tong Zhao, and Meng Jiang. 2020a. Identifying referential intention with heterogeneous contexts. In *Proceedings of The Web Conference 2020*, pages 962–972.

Wenhao Yu, Chenguang Zhu, Yuwei Fang, Donghan Yu, Shuohang Wang, Yichong Xu, Michael Zeng, and Meng Jiang. 2021a. Dict-bert: Enhancing language model pre-training with dictionary. *arXiv preprint arXiv:2110.06490*.

Wenhao Yu, Chenguang Zhu, Zaitang Li, Zhiting Hu, Qingyun Wang, Heng Ji, and Meng Jiang. 2020b. A survey of knowledge-enhanced text generation. *ACM Computing Survey (CSUR)*.

Wenhao Yu, Chenguang Zhu, Tong Zhao, Zhichun Guo, and Meng Jiang. 2021b. Sentence-permuted paragraph generation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Qingkai Zeng, Jinfeng Lin, Wenhao Yu, Jane Cleland-Huang, and Meng Jiang. 2021. Enhancing taxonomy completion with concept generation via fusing relational representations. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*.

Hongming Zhang, Xin Liu, Haojie Pan, Hao Ke, Jiefu Ou, Tianqing Fang, and Yangqiu Song. 2021. Aser: Towards large-scale commonsense knowledge acquisition via higher-order selectional preference over eventualities. *ArXiv*, abs/2104.02137.

Houyu Zhang, Zhenghao Liu, Chenyan Xiong, and Zhiyuan Liu. 2020a. Grounded conversation generation as guided traverses in commonsense knowledge graphs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2031–2043, Online. Association for Computational Linguistics.

Houyu Zhang, Zhenghao Liu, Chenyan Xiong, and Zhiyuan Liu. 2020b. Grounded conversation generation as guided traverses in commonsense knowledge graphs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2031–2043, Online. Association for Computational Linguistics.

Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. 2019. ERNIE: Enhanced language representation with informative entities. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1441–1451, Florence, Italy. Association for Computational Linguistics.

Peixiang Zhong, Di Wang, Pengfei Li, Chen Zhang, Hao Wang, and C. Miao. 2021. Care: Commonsense-aware emotional response generation with latent concepts. In *AAAI*.

Hao Zhou, Tom Young, Minlie Huang, Haizhou Zhao, Jingfang Xu, and Xiaoyan Zhu. 2018. Commonsense knowledge aware conversation generation with graph attention. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pages 4623–4629. ijcai.org.

Pei Zhou, Rahul Khanna, Seyeon Lee, Bill Yuchen Lin, Daniel Ho, Jay Pujara, and Xiang Ren. 2020. Rica: Evaluating robust inference capabilities based on commonsense axioms. *arXiv: Computation and Language*.