Learning from Task Instructions

Wenpeng Yin
Temple University
Philadelphia, USA
wenpeng.yin@temple.edu

Hinrich Schütze
LMU Munich
Munich, Germany
hinrich@hotmail.com

Abstract

The progress of natural language processing (NLP) is primarily driven by machine learning that optimizes a system on a large-scale set of task-specific labeled examples. This learning paradigm limits the ability of machines to have the same capabilities as humans in handling new tasks since humans can often solve unseen tasks with a couple of examples accompanied by task instruction. In addition, we may not have a chance to prepare task-specific examples of large-volume for new tasks because we cannot foresee what task needs to be addressed next and how complex to annotate for it. Therefore, task instructions act as a novel and promising resource for supervision.

This tutorial targets researchers and practitioners who are interested in AI and ML technologies for NLP generalization in a low-shot scenario. In particular, we will present a diverse thread of instruction-driven NLP studies that try to answer the following questions: (i) What is task instruction? (ii) How to construct task instructions? (iii) How to encode task instruction? (iv) How generalizable are the systems trained on task instructions? (v) How robust is learning from task instructions? We will discuss several lines of frontier research that tackle those challenges and will conclude the tutorial by outlining directions for further investigation.

1 Introduction

This proposal is driven by a fundamental question of task generalization in NLP: how to comprehend a new task if labeled examples are pretty limited? One goal of AI is to build a system that can continually understand and solve new tasks. Labeled examples, as the mainstream task representation, are unlikely to be available in large numbers or even do not exist. Then, is there any other task representation that can contribute to task comprehension? Task instructions provide another dimension of supervision for expressing the task semantics. Instructions often contain more abstract and more comprehensive knowledge of the target task than individual labeled examples. With the availability of task instructions, systems can be quickly built to handle new tasks, especially when task-specific annotations are scarce (Wang et al., 2022; Yin et al., 2022). Instruction Learning is inspired by the typical human learning for new tasks, e.g., a little kid can well solve a new mathematical task by learning from its instruction and a few examples. The communities of machine learning and NLP, however, have paid little attention to this new learning paradigm.

Despite the importance, frontier research in instruction learning is still struggling with the following questions. First, should instructions be constructed to express the target task as detailed as possible (e.g., MTurk instructions (Mishra et al., 2022)) or to align with the format of supervising tasks (e.g., textual entailment (Yin et al., 2019) or language modeling (Brown et al., 2020a)) as well as possible? Second, how to effectively encode instructions that may consist of some specific requirements such as "maximal output length 5", and "do not generate anything else apart from one of the following $\cdots$"? Third, how generalizable are the systems trained on task instructions in dealing with unseen tasks? Last, how robust will a pretrained instruction-driven system be to cope with instructions that are written by different annotators or varying degrees of abstraction?

In this tutorial, we will systematically review several lines of frontier research on developing systems that are supervised by task instructions. Beyond introducing pioneering work that parsed instructions to cope with individual tasks, such as soccer game (Kuhlmann et al., 2004), software control (Branavan et al., 2009, 2011), etc., we will focus on recent approaches for cross-task generalization given task instructions. Specifically, three dimensions of instructions will be introduced: (i) entailment-oriented task instructions, where tasks are converted into textual entailment and instructions are constructed to fit the indirect supervision of textual entailment; (ii) PLM-oriented task instructions (i.e., prompts), where NLP tasks are transformed into language modeling problems and prompts act as instructions to probe the pretrained language model (PLM); (iii) human-oriented task instructions (i.e, Amazon MTurk instructions), where instructions are created by users who are not expert of machine learning or NLP but machines are expected to understand them.

Participants will learn about recent trends and emerging challenges in this topic, representative tools and learning resources to obtain ready-to-use models, and how related technologies benefit end-user NLP applications.

2 Outline of Tutorial Content

This **half-day** tutorial presents a systematic overview of recent advancements in NLP with supervision from task instructions. The detailed contents are outlined below.

2.1 Background and motivation [15min]

We will define the main research problem and motivate the topic by presenting several real-world NLP and knowledge-driven AI applications, as well as several key challenges that are at the core of conventional machine learning.

2.2 Following instructions for particular tasks [30min]

Early works interpreted instructions through semantic parsing in the context of providing natural language interfaces to computer systems. We first review some **non-NLP** work, such as following navigational instructions (Vogel and Jurafsky, 2010; Chen and Mooney, 2011; Tellex et al., 2011; Chen, 2012; Kim and Mooney, 2012; Artzi and Zettlemoyer, 2013), software control (Branavan et al., 2009, 2010), querying databases (Clarke et al., 2010), understanding visual scenes in the physical world (Matuszek et al., 2012; Krishnamurthy and Kollar, 2013; Srivastava et al., 2017a), VirtualHome environment (Puig et al., 2018) and playing games based on text (Chen and Mooney, 2008; Eisenstein et al., 2009; Liang et al., 2009; Branavan et al., 2011; Goldwasser and Roth, 2011; Bisk et al., 2016), then focus on the **NLP** domain. For example, Srivastava et al. (2018) studied zero-shot email classification by parsing the NL quantification. Some works parsed the task explanations to learn new email categories (Srivastava et al., 2017b), or to generate noisy labeled datasets for training classifiers in relation extraction and machine comprehension (Hancock et al., 2018; Ye et al., 2020; Wang et al., 2020).

Whether they are non-NLP or NLP-specific, this subsection focuses on a single task by learning a task-specific instruction interpreter.

2.3 Entailment-oriented task instructions [30min]

For most zero/few-shot text classification tasks, such as topic classification, entity typing, relation identification, etc., the main obstacle is to let systems understand the semantics of labels. In contrast to conventional supervised classifiers, which converted labels into indices, a textual entailment based methodology takes into account the input semantics as well as label semantics. In specific, we will introduce typical work that treats different topics (Yin et al., 2019), stances (Xu et al., 2022), entity types (Li et al., 2022), and entity relations (Xia et al., 2021; Sainz et al., 2021, 2022) as hypotheses (i.e., instructions) and the inputs as premises, then makes use of the indirect supervision from textual entailment to handle a variety of classification tasks with open-domain texts and open-form labels.

2.4 PLM-oriented task instructions [30min]

Prompting is the practice of representing a task as a brief utterance in order to query a PLM for a response. PLM-oriented instruction learning is able to get rid of human-annotated supervision (e.g., textual entailment) and relies on fully unsupervised language models. We will briefly review literature that employed prompts for sentence-level tasks, such as machine translation, question answering (Radford et al., 2019), sentiment analysis, textual entailment (Schick and Schütze, 2020, 2021a), etc., and mainly elaborate on PET (Schick and Schütze, 2020, 2021a, 2022), which makes use of prompts for real-world few-shot NLP.

In addition, the process of prompt engineering is critical for successful deployment as choices in prompting can affect downstream predictions significantly. It motivates a practical challenge: how can users create, refine, and share prompts? We will introduce (Bach et al., 2022) that created a Web-based UI, called "*PromptSource*", that enables developers to write prompts in a templating language and immediately view their outputs on different examples, based on which over 2,000 open-source prompts have been collected for roughly 170 NLP tasks. This collection, named "*Public Pool of Prompts (P3)*" has allowed users to check zero-shot cross-task generalization (Sanh et al., 2021), zero- and few-shot cross-lingual generalization (Lin et al., 2021), and in-context learning (Min et al., 2021). *While these prompt-based results are encouraging, such prompts are often too simplistic, whereas many real NLP problems cannot be effectively formulated as short prompts.*

2.5 Human-oriented task instructions [30min]

To handle the limitations of prompt-based PLMs, we further introduce instructions that are human-oriented. Two reasons: (i) prompts are too short to express the details of a task; (ii) prompts are PLM-oriented, while in the real world, we hope our AI system can be operated by humans who are not machine learning experts but can instruct what and how to do.

We first introduce (Efrat and Levy, 2020) that tested GPT-2 (Radford et al., 2019) to understand real-world MTurk instructions to annotate some popular datasets, and concluded that GPT-2 works poorly, then highlight (Mishra et al., 2021; Wang et al., 2022) that collected more than 1.6k cross-lingual NLP tasks with MTurk instructions consisting of items like `title`, `definition`, `things to avoid`, etc., and claimed that BART (Lewis et al., 2020) and GPT-3 (Brown et al., 2020b) benefit from instructions to generalize across tasks, and finally mention how Yin et al. (2022) built machines to learn tasks incrementally with MTurk instructions.

2.6 Robustness of learning from task instructions [30min]

Most instruction-driven systems assume that each task has a single instruction. We can imagine that different users can convey a task with instructions of distinct textual expressions. Some prompt-based PLMs also show varying performance in dealing with prompts of different templates (Schick and Schütze, 2021b; Kojima et al., 2022). Questions arise: are the PLMs robust enough to handle expression-varying instructions of the same task? To the end, we will introduce the work by Gu and Yin (2022) that explored the robustness of pretrained instruction learning system in handling (i) the same task with distinct instructions written by different MTurkers, and (ii) instruction of varying-degrees of abstractions.

2.7 Future research directions [15min]

In this section, we will discuss future work in the following threads: (i) generating instructions from labeled examples, (ii) explainable instruction learning, and (iii) how to encode instructions without the help of labeled examples, etc.

3 Specification of the Tutorial

The proposed tutorial is considered a **cutting-edge** tutorial that introduces new frontiers in NLP/AI research. The presented topic has not been covered by ACL/EMNLP/NAACL/EACL/COLING tutorials in the past 4 years.

Audience and Prerequisites Based on the level of interest in this topic, we expect around 100 participants. While no specific background knowledge is assumed of the audience, it would be the best for the attendees to know about basic deep learning technologies, pre-trained language models (e.g. BERT).

4 Tutorial Instructors

The following are biographies of the speaker.

Wenpeng Yin is an Assistant Professor at Temple University. Prior to joining Temple, he was a Senior Research Scientist at Salesforce Research (8/2019-12/2021), a postdoctoral researcher at UPenn (10/2017-7/2019), and got his Ph.D. degree from the LMU Munich, Germany, in 2017. Dr. Yin's research focuses on natural language processing with three sub-areas: (i) learning from task instructions; (ii) information extraction; (iii) learning with limited supervision. Additional information is available at `www.wenpengyin.org`.

Hinrich Schütze is Chair of Computational Linguistics and co-director of the Center of Information and Language Processing at Ludwig-Maximilians-Universität München. Prior to joining LMU Munich, he was a Professor of Theoretical Computational Linguistics at the University of Stuttgart. Hinrich holds a PhD in computational linguistics from Stanford University. We was a scientist at the XEROX Palo Alto Research Center from 1995 to 2000 and was involved in leading roles in a number of Silicon Valley startups from 2000 to 2004. Additional information is available at `https://schuetze.cis.lmu.de`.

References

Yoav Artzi and Luke Zettlemoyer. 2013. Weakly supervised learning of semantic parsers for mapping instructions to actions. *Trans. Assoc. Comput. Linguistics*, 1:49–62.

Stephen H. Bach, Victor Sanh, Zheng Xin Yong, Albert Webson, Colin Raffel, Nihal V. Nayak, Abheesht Sharma, Taewoon Kim, M. Saiful Bari, Thibault Févry, Zaid Alyafeai, Manan Dey, Andrea Santilli, Zhiqing Sun, Srulik Ben-David, Canwen Xu, Gunjan Chhablani, Han Wang, Jason Alan Fries, Maged Saeed AlShaibani, Shanya Sharma, Urmish Thakker, Khalid Almubarak, Xiangru Tang, Dragomir R. Radev, Mike Tian-Jian Jiang, and Alexander M. Rush. 2022. Promptsource: An integrated development environment and repository for natural language prompts. In *Proceedings of ACL System Demonstrations*, pages 93–104.

Yonatan Bisk, Deniz Yuret, and Daniel Marcu. 2016. Natural language communication with robots. In *Proceedings of NAACL*, pages 751–761.

S. R. K. Branavan, Harr Chen, Luke Zettlemoyer, and Regina Barzilay. 2009. Reinforcement learning for mapping instructions to actions. In *Proceedings of ACL*, pages 82–90.

S. R. K. Branavan, David Silver, and Regina Barzilay. 2011. Learning to win by reading manuals in a monte-carlo framework. In *Proceedings of ACL*, pages 268–277.

S. R. K. Branavan, Luke Zettlemoyer, and Regina Barzilay. 2010. Reading between the lines: Learning to map high-level instructions to commands. In *Proceedings of ACL*, pages 1268–1277.

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. 2020a. 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*.

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. 2020b. Language models are few-shot learners. In *Proceedings of NeurIPS*.

David L. Chen. 2012. Fast online lexicon learning for grounded language acquisition. In *Proceedings of ACL*, pages 430–439.

David L. Chen and Raymond J. Mooney. 2008. Learning to sportscast: a test of grounded language acquisition. In *Proceedings of ICML*, volume 307, pages 128–135.

David L. Chen and Raymond J. Mooney. 2011. Learning to interpret natural language navigation instructions from observations. In *Proceedings of AAAI*, pages 859–865.

James Clarke, Dan Goldwasser, Ming-Wei Chang, and Dan Roth. 2010. Driving semantic parsing from the world's response. In *Proceedings of CoNLL*, pages 18–27.

Avia Efrat and Omer Levy. 2020. The turking test: Can language models understand instructions? *CoRR*, abs/2010.11982.

Jacob Eisenstein, James Clarke, Dan Goldwasser, and Dan Roth. 2009. Reading to learn: Constructing features from semantic abstracts. In *Proceedings of EMNLP*, pages 958–967.

Dan Goldwasser and Dan Roth. 2011. Learning from natural instructions. In *Proceedings of IJCAI*, pages 1794–1800.

Jiasheng Gu and Wenpengg Yin. 2022. Robustness of learning natural language instructions. *CoRR*, abs/2205.11352.

Braden Hancock, Paroma Varma, Stephanie Wang, Martin Bringmann, Percy Liang, and Christopher Ré. 2018. Training classifiers with natural language explanations. In *Proceedings of ACL*, pages 1884–1895.

Joohyun Kim and Raymond J. Mooney. 2012. Unsupervised PCFG induction for grounded language learning with highly ambiguous supervision. In *Proceedings of EMNLP*, pages 433–444.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *CoRR*, abs/2205.11916.

Jayant Krishnamurthy and Thomas Kollar. 2013. Jointly learning to parse and perceive: Connecting natural language to the physical world. *Trans. Assoc. Comput. Linguistics*, 1:193–206.

Gregory Kuhlmann, Peter Stone, Raymond Mooney, and Jude Shavlik. 2004. Guiding a reinforcement learner with natural language advice: Initial results in robocup soccer. In *The AAAI workshop on supervisory control of learning and adaptive systems*, pages 30–35.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of ACL*, pages 7871–7880.

Bangzheng Li, Wenpeng Yin, and Muhao Chen. 2022. Ultra-fine entity typing with indirect supervision from natural language inference. *Trans. Assoc. Comput. Linguistics*, 10:607–622.

Percy Liang, Michael I. Jordan, and Dan Klein. 2009. Learning semantic correspondences with less supervision. In *Proceedings of ACL*, pages 91–99.

Xi Victoria Lin, Todor Mihaylov, Mikel Artetxe, Tianlu Wang, Shuohui Chen, Daniel Simig, Myle Ott, Naman Goyal, Shruti Bhosale, Jingfei Du, Ramakanth Pasunuru, Sam Shleifer, Punit Singh Koura, Vishrav Chaudhary, Brian O'Horo, Jeff Wang, Luke Zettlemoyer, Zornitsa Kozareva, Mona T. Diab, Veselin Stoyanov, and Xian Li. 2021. Few-shot learning with multilingual language models. *CoRR*, abs/2112.10668.

Cynthia Matuszek, Nicholas FitzGerald, Luke Zettlemoyer, Liefeng Bo, and Dieter Fox. 2012. A joint model of language and perception for grounded attribute learning. In *Proceedings of ICML*.

Sewon Min, Mike Lewis, Luke Zettlemoyer, and Hannaneh Hajishirzi. 2021. Metaicl: Learning to learn in context. *CoRR*, abs/2110.15943.

Swaroop Mishra, Daniel Khashabi, Chitta Baral, and Hannaneh Hajishirzi. 2021. Cross-task generalization via natural language crowdsourcing instructions. *CoRR*, abs/2104.08773.

Swaroop Mishra, Daniel Khashabi, Chitta Baral, and Hannaneh Hajishirzi. 2022. Cross-task generalization via natural language crowdsourcing instructions. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 3470–3487. Association for Computational Linguistics.

Xavier Puig, Kevin Ra, Marko Boben, Jiaman Li, Tingwu Wang, Sanja Fidler, and Antonio Torralba. 2018. Virtualhome: Simulating household activities via programs. In *Proceedings of CVPR*, pages 8494–8502.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI blog*.

Oscar Sainz, Oier Lopez de Lacalle, Gorka Labaka, Ander Barrena, and Eneko Agirre. 2021. Label verbalization and entailment for effective zero and few-shot relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 1199–1212.

Oscar Sainz, Itziar Gonzalez-Dios, Oier Lopez de Lacalle, Bonan Min, and Eneko Agirre. 2022. Textual entailment for event argument extraction: Zero- and few-shot with multi-source learning. In *Findings of the Association for Computational Linguistics: NAACL 2022, Seattle, WA, United States, July 10-15, 2022*, pages 2439–2455.

Victor Sanh, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, Manan Dey, M. Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal V. Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Févry, Jason Alan Fries, Ryan Teehan, Stella Biderman, Leo Gao, Tali Bers, Thomas Wolf, and Alexander M. Rush. 2021. Multitask prompted training enables zero-shot task generalization. *CoRR*, abs/2110.08207.

Timo Schick and Hinrich Schütze. 2020. Few-shot text generation with pattern-exploiting training. *CoRR*, abs/2012.11926.

Timo Schick and Hinrich Schütze. 2021a. Exploiting cloze-questions for few-shot text classification and natural language inference. In *Proceedings of EACL*, pages 255–269.

Timo Schick and Hinrich Schütze. 2021b. True few-shot learning with prompts - A real-world perspective. *CoRR*, abs/2111.13440.

Timo Schick and Hinrich Schütze. 2022. True few-shot learning with prompts - A real-world perspective. *Trans. Assoc. Comput. Linguistics*, 10:716–731.

Shashank Srivastava, Igor Labutov, and Tom Mitchell. 2017a. Learning classifiers from declarative language. In *NIPS Workshop on Learning from Limited Data*.

Shashank Srivastava, Igor Labutov, and Tom M. Mitchell. 2017b. Joint concept learning and semantic parsing from natural language explanations. In *Proceedings of EMNLP*, pages 1527–1536.

Shashank Srivastava, Igor Labutov, and Tom M. Mitchell. 2018. Zero-shot learning of classifiers from natural language quantification. In *Proceedings of ACL*, pages 306–316.

Stefanie Tellex, Thomas Kollar, Steven Dickerson, Matthew R. Walter, Ashis Gopal Banerjee, Seth J. Teller, and Nicholas Roy. 2011. Approaching the symbol grounding problem with probabilistic graphical models. *AI Mag.*, 32(4):64–76.

Adam Vogel and Daniel Jurafsky. 2010. Learning to follow navigational directions. In *Proceedings of ACL*, pages 806–814.

Yizhong Wang, Swaroop Mishra, Pegah Alipoormolabashi, Yeganeh Kordi, Amirreza Mirzaei, Anjana Arunkumar, Arjun Ashok, Arut Selvan Dhanasekaran, Atharva Naik, David Stap, Eshaan Pathak, Giannis Karamanolakis, Haizhi Gary Lai, Ishan Purohit, Ishani Mondal, Jacob Anderson, Kirby Kuznia, Krima Doshi, Maitreya Patel, Kuntal Kumar Pal, Mehrad Moradshahi, Mihir Parmar, Mirali Purohit, Neeraj Varshney, Phani Rohitha Kaza, Pulkit Verma, Ravsehaj Singh Puri, Rushang Karia, Shailaja Keyur Sampat, Savan Doshi, Siddhartha Mishra, Sujan Reddy, Sumanta Patro, Tanay Dixit, Xudong Shen, Chitta Baral, Yejin Choi, Hannaneh Hajishirzi, Noah A. Smith, and Daniel Khashabi. 2022. Benchmarking generalization via in-context instructions on 1, 600+ language tasks. *CoRR*, abs/2204.07705.

Ziqi Wang, Yujia Qin, Wenxuan Zhou, Jun Yan, Qinyuan Ye, Leonardo Neves, Zhiyuan Liu, and Xiang Ren. 2020. Learning from explanations with neural execution tree. In *Proceedings of ICLR*.

Congying Xia, Wenpeng Yin, Yihao Feng, and Philip S. Yu. 2021. Incremental few-shot text classification with multi-round new classes: Formulation, dataset and system. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 1351–1360.

Hanzi Xu, Slobodan Vucetic, and Wenpeng Yin. 2022. Openstance: Real-world zero-shot stance detection. In *Proceedings of CoNLL*.

Qinyuan Ye, Xiao Huang, Elizabeth Boschee, and Xiang Ren. 2020. Teaching machine comprehension with compositional explanations. In *Proceedings of EMNLP Findings*, volume EMNLP 2020 of *Findings of ACL*, pages 1599–1615.

Wenpeng Yin, Jamaal Hay, and Dan Roth. 2019. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. 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 2019, Hong Kong, China, November 3-7, 2019*, pages 3912–3921. Association for Computational Linguistics.

Wenpeng Yin, Jia Li, and Caiming Xiong. 2022. ConTinTin: Continual learning from task instructions. In *Proceedings of ACL*, pages 3062–3072.