Robust, Generalizable and Minimally Supervised Knowledge Acquisition from Human Language

Muhao Chen
Assistant Research Professor of Computer Science, USC

1 Introduction and Background

My research seeks to advance intelligent computational systems by developing the tools and techniques required for acquiring and representing the world knowledge, and achieving synergetic knowledge-sharing between systems, people and environments. Therefore, my long-term goal is to leverage minimal supervision to allow machines to confidently understand relations involving concepts, entities and events in human languages, as well as interactions between objects in nature (such as molecules and biomolecules). In the near term, I am motivated by the objective of developing new technologies in representation learning and information extraction, and extending their use in various tasks used for natural language processing, knowledge base construction, computational biology and medicine.

In this decade, the power of AI systems in various application domains was greatly enhanced by representation learning technologies for automatically discovering and acquiring relations, patterns and properties of objects from large-scale data. In particular, such technologies involve information extraction, language modeling and constrained learning. My research seeks to further advance these technologies, by focusing on knowledge acquisition from data under different modalities, and considering both with and without plausible-supervision-signals scenarios. My investigations in these research areas produced over 60 publications in leading AI, NLP and machine learning venues, documenting novel techniques that were deployed in many real-world applications, and delivered significant benefits in various computational and interdisciplinary areas. In the rest of this statement, I will briefly summarize my recent investigations and directions, illustrated in the research roadmap shown in Fig. 1. Then I will outline some of the exciting directions I plan to pursue in my future research.

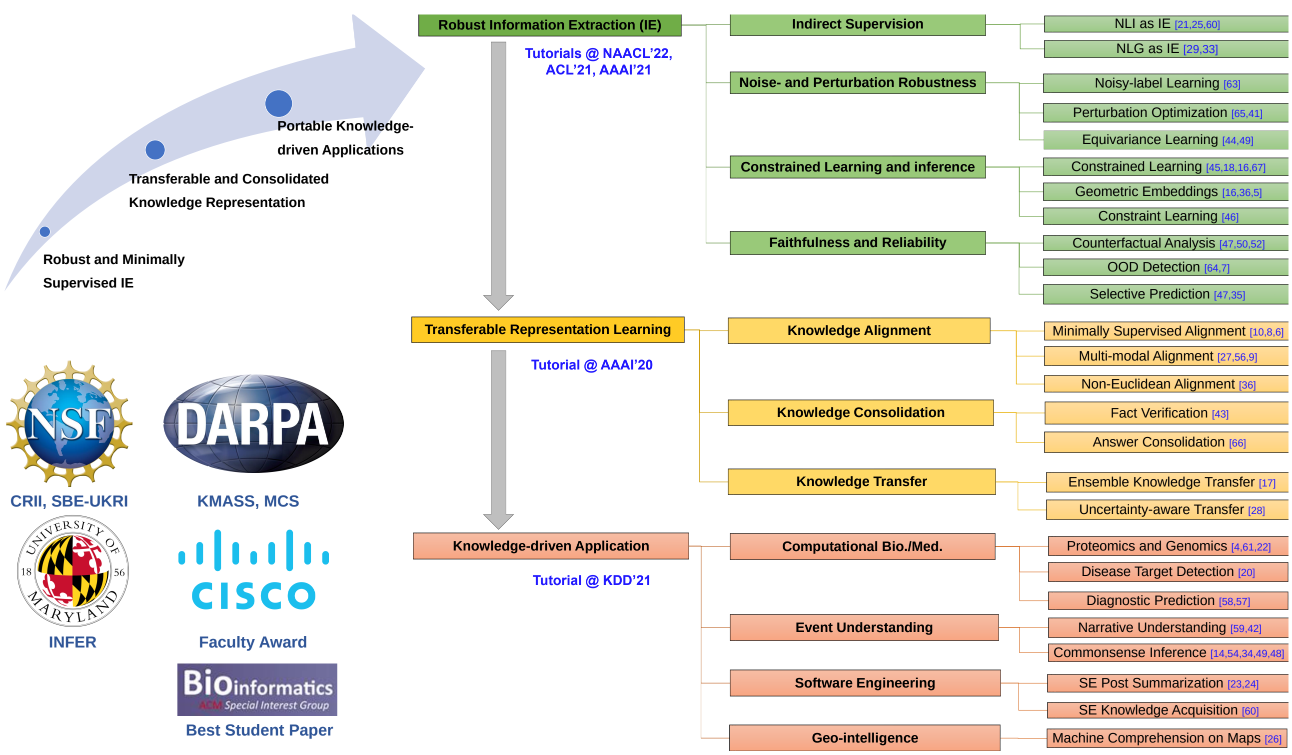

Figure 1: A roadmap of my research on robust, generalizable and minimally supervised knowledge acquisition.

Robust and Minimally Supervised Knowledge Extraction Structural knowledge representations, such as knowledge graphs (KGs), represent an integral part of intelligent systems in nearly all domains. Despite their importance, existing efforts for acquiring structural knowledge representation are generally conducted manually. This leads to the cost of tens of millions of dollars to produce just one mid-scale commodity KG in the general domain, and this cost can be significantly higher in scientific domains like biology and medicine. To alleviate these costly efforts and provide reliable systems for automatically acquiring knowledge, my research has focused on *robust* and *minimally supervised* learning and inference technologies for knowledge extraction from natural language text. Specifically, my research systematically leads to transformative advancements in the following four dimensions.

(1) Indirect supervision. Learning to extract structural knowledge has largely relied on *direct supervision* from structurally annotated corpora that are similarly expensive as a structural knowledge representation itself [32]. I instead investigate a novel direction of *indirect supervision*, leading to robust and generalizable knowledge extraction models without sole reliance on task-specific annotations. In particular, my study has produced principled approaches for reformulating and transferring supervision signals from natural language inference (NLI) [25], summarization [29] and linguistic pattern matching [33]. This reformulation allows rich (indirect) supervision signals to be transferred from well-developed learning resources and models for signal-providing tasks that align well with knowledge extraction. It also emancipated knowledge extraction from the limitation of fixed label sets, allowing the inference of new types of knowledge that were unseen in training. In this context, I have also explored with semantic representation of tasks and labels [21, 14] to further reduce the need of direct task supervision. This systematic study of indirectly supervised learning has led to SOTA performance on a large number of benchmarks for relation extraction, named entity recognition, ultra-fine entity typing, event extraction and event process typing. Specifically, for all those tasks, my systems have demonstrated excellence in extremely few-shot [29, 21, 5] or zero-shot performances [25] that were close to those that perviously offered by full-shot, directly supervised models.

(2) Noise- and perturbation-robustness. In addition to insufficiency of annotated data, the cost and difficulty of structural annotation often lead to significant training noise. In the same context, real-world application scenarios often expose the model with way larger and more diverse data, for which the inference of model needs to frequently handle perturbations and out-of-distribution (OOD) exceptions. My study accordingly enhance the robustness of the model from two perspective. Towards robust training, my study developed a co-regularized knowledge distillation approach that can proactively identifying noisy training instances and preventing the discriminative model from fitting the noise [63]. This leads to significant improvement in both noise-robustness and computational efficiency over previous ensemble-based denoising and noise-filtering methods. In this context, my study also proposed sharpness-aware minimization with dynamic reweighting (δ-SAM [65]) to further enhance the model robustness using adversarial perturbation training, as well as self-supervised cross-lingual perturbation training [41]. On the other hand, to enhance the robustness in inference, I have studied margin-based contrastive learning methods [64, 7] that led to near-perfect unsupervised OOD detection performance, helping the model selectively identify cases where no extraction should be made. I also developed structure-aware equivariance learning techniques [44, 49] to allow data-to-text generation models to generate consistent representation for structural priors where semantic-invariant perturbations are free to be introduced. Those technologies systematically improves the reliability of knowledge extraction systems in real-world scenarios where training and inference phases are abundant with noise, perturbations and exceptions.

(3) Logically constrained learning and inference. Extracts are not standalone and can possess complex logical dependencies. A robust knowledge extraction system needs to ensure that the extracts are self-contained, and free of inconsistency and redundancy. My work accordingly suggests solutions to this problem with novel constrained learning and inference approaches. Specifically, I have studied joint constrained learning approaches for enforcing logical consistency in relation extraction tasks [45], probabilistic constrained learning with t-norm based optimization [18], logically constrained learning for linear relational embeddings [11] and probablistic box embeddings [16]. Considering that logical constraints may be costly to define and hard to articulate, my recent study also proposed the approach to learn linear inequalities for automatically capturing logical constraints

from data [46].

(4) Faithfulness. Current knowledge extraction models are mainly developed on large pre-trained language models and are short of training annotations in general. In this situation, my study has discovered that pre-training knowledge, distribution biases or existing annotation artifacts could often cause models to unfaithfully extract what is described in a given context, but instead to "guess" with a context-irrelevant extract using prediction shortcuts [50, 52]. Faithfulness, while being an under-explored research area, is undoubtedly a premise of reliable information extraction. In this context, my study has so far delivered several pilot studies to mitigate prediction shortcuts in entity-centric and event-centric information extraction with counterfactual analysis [50, 47], and counterfactual data augmentation [52]. On the other hand, to ensure that models make selective decisions on exception cases where nothing should be extracted, we contribute with selective prediction techniques based on high-order metric learning [28, 35] and Dirichlet parameterization [47].

My main in this line of research were summarized in the tutorials I presented at NAACL 2022 [3], ACL 2021 [12] and other invited talks, and led to the support I am receiving from the DARPA KMASS program, the DARPA MCS program, and the Cisco Faculty Research Award.

Transferable Representation Learning for Structural Knowledge. Structural representation learning is the requisite for incorporating symbolic knowledge into deep learning models. A key contribution I have made to this field is on the *transferability* of such representations. Different domains or sources of data, or even different languages, often provide *interchangeable* and *complementary* knowledge. Hence, it is particularly important to develop a universal representation learning method that captures the association of knowledge across multiple data sources with minimal supervision, and support with credible knowledge transfer across different domains. I started this line of research and provided the first embedding framework that bridges multiple language-specific KGs [10, 15], by performing semi-supervised alignment of multiple relational embedding models. To more precisely capture the knowledge association with minimal supervision, I have extensively extended the alignment learning process based on iterative co-training [8], multi-view representation [9, 56], incidental supervision from free text [6], unsupervised visual pivoting [27] and coarse-to-fine entity linking [19]. I have also devised relational embedding techniques that are robust against scarcity and structural heterogeneity of data, using techniques based on box embeddings [16], concept contextualization [37] and attentive neighborhood aggregation [38]. Particularly, for highly complex knowledge-representation structures, I devised on new paradigms for non-linear embedding spaces [36, 30, 5]. For knowledge transfer from multiple sources of (inconsistent) learning resources, my work addresses the problem of inducing trustworthy inference results with ensemble knowledge transfer [17]. In this context, my study also contributes with answer consolidation [66] and multi-modal fact verification techniques [43] that help resolve the redundancy and inconsistency of local extracts for global knowledge representation.

This line of research has received a wide recognition by the community, and the importance of this contribution has been recognized by over a thousand citations in the past four years. A wide spectrum of applications have also been benefited from the techniques proposed in my papers and follow-up works. The advancement in this research topic has been featured in my tutorial at AAAI-2020 [2] and our recent survey paper [39], and has been recognized with an NSF CRII Award in 2021.

Portable Knowledge-driven Applications The robust knowledge extraction and knowledge transfer technologies allow efficient and broad utility of the knowledge in computational research. This allows my study to further enhance the various narrative understanding [59, 42] and commonsense reasoning tasks [54, 14, 34, 33, 49, 48] that are at the core of current NLP research. My recent study also extends the utility of robust information extraction [60] and summarization techniques [23, 24] for knowledge acquisition and dissemination from online programming tutorials and discussions, aiming at helping software developers in making informed decisions and supporting the development of knowledge-based tools. Besides, I also gained critical experience on how to best transfer the above two lines of technologies to tackle important tasks in computational biology and medicine, including protein-protein interaction prediction [4], proteomic mutation effect estimation [61], circular RNA detection [22], disease target identification [20], and clinical diagnostic prediction [57, 58]. In addition, a recent thread of my research also interfaces language modeling with geographical data [26], aiming at benefiting automated machine comprehension and dissemination of digital map information. I am excited about the broad utility of knowledge acquired from my fundamental research on information extraction and knowledge transfer, and will continue the investigation on solving important problems in various related areas.

2 Research Agenda: Directions for Future Work

My future research will continue to focus on improving machines' ability of efficiently and reliably acquiring general-domain and expert knowledge with minimal supervision, and leverage transferable representations to solve problems in various domains and interdisciplinary areas. Particularly, I intend to extend my investigation in the following directions.

Event-Centric NLP. Human languages evolve to communicate about events happening in the real world. Therefore, understanding events plays a critical role in natural language understanding. A key challenge in accomplishing this mission lies in the fact that events are not just simple, standalone predicates. Rather, they are often described at different granularities, temporally form procedures, and are always directed by specific goals and sub-goals. Understanding events requires the understanding of how events are connected, form procedural or membership structures, and recognizing typical properties of events (e.g., space, time, salience, essentiality, implicitness, preconditions, consequences, etc.). Our prior studies have focused on inferring the logical constraints [45, 46], analogical properties [54], evolution patterns [67], salience [59] and membership relations [14] of events. Yet, more fundamental challenges persist. Particularly, unlike entities, events have an internal structure, whereby ensuring consistent information extraction for internal components and external relations of events remain an unresolved problem. Besides, machine comprehension of causal relations, conditions and consequences requires unrealized human-like cognitive understanding. As events may evolve and be described in different ways in different documents, inferring the relations of events across documents, consolidating highly diverse cross-document event descriptions into unique and consistent global knowledge representation represents a daunting challenge to event-centric information extraction. Moreover, reasoning about events challenges commonsense reasoning agents with common implicitness of event mentions, event arguments and event properties (e.g., essentiality [55], preconditions [33], and spatial attributes). My recent tutorials at ACL 2021 [12] and AAAI 2021 [13] have systematically summarized the current (pre-mature) status of event-centric NLP, and positioned the emerging fundamental research challenges, including the above, that will be at the core of my future research. In this context, I also plan to investigate the use of event knowledge to improve the coherence of narrative (e.g., our prior study [59]) and dialogue systems, enhance the factuality of summarization systems, as well as realize clinical event understanding to tackle important but expensive clinical diagnostic tasks [58, 57].

Harnessing Massive Language Models. Massive language models (also known as foundation models) such as GPT-3 and PaLM have excelled in many NLP tasks with their strong ability of distilling knowledge from Web-scale pre-training corpora, and raised opportunities in research directions such as in-context learning. Despite the progress, there are two critical issues with NLP solutions based on massive language models. First, massive language models still fall short of supporting reasoning, including logical, quantitative and cognitive reasoning. Our recent study on machine comprehension of medical reports and tabular data has proven that massive language models still fail to correctly infer the relations between time intervals, and do not meaningfully support numeracy. Our study also found that GPT-3 fails at many cognitive understanding tasks. For example, they achieve an AUC of 63% when inferring the essentiality of events [55], whereas humans can achieve around 87%. Second, after costing tens of millions of dollar to be trained, massive language models have to remain fixed within their year-long life cycles, causing them to be inadaptive to vastly streaming new information about the ever-changing world. My future research will tackle these issues in two directions. First, I will develop various mid-scale language models where dedicated kernel functions or neural symbolic modules are incorporated into the Transformer architectures, seeking to realize diverse types of reasoning processes. These models will form mixture-of-experts (MoE) or hero-gang structures together with the massive language model, providing complementary reasoning abilities. Second, towards adaptation to changing information, I will mainly investigate two methodologies. Specifically, I will continue our study on parameter-efficient adaptation [53] to allow plug-in of memory modules about new knowledge into the MoE structure, for which added information in the memory modules may come from both new data and human-in-the-loop. Moreover, I will leverage our robust information extraction technologies to timely capture new information from the Web, and realize retrieval-augmentation of massive language models at inference.

Generalizing Indirect Supervision for Scalable NLP. My prior studies have demonstrated the success of indirect supervision from NLI and NLG tasks in enhancing information extraction within the scope of single documents [25, 21, 29]. I will further extend this line of study in three directions. First, to allow more enriched knowledge extraction, it is essential to enable the extraction process across documents. As cross-document extraction tasks, such as cross-document co-reference resolution and cross-document relation extraction suffer more severely from insufficient training data than their single-document counterparts, my study will focus on (i) multi-hop dense retrieval approaches [51] for discovering evidence that supports cross-document relations, and (ii) indirect supervision from multi-document tasks, such as multi-document summarization and sentence fusion. Second, beyond the use of existing source tasks for supervision signals, I will also design linguistic pattern mining and generative data augmentation methods that can automatically find or generate large-scale weak supervision data (a preliminary study has been conducted for preconditioned commonsense inference [33]). In the same context, I will further study methods that proactively select and filter task-specific weak supervision data following my prior studies on unsupervised denoising [63, 65]. Last, to understand the learnability of different sources of indirect supervision, we will study methods to quantify the contribution made by each source task to the target task objectives, and also indentify a parameter-level affinity measurement that helps select the substructures of the models for optimized cross-task signal transfer.

Equivariance Learning in NLP As an important but largely under-explored component of robust NLP systems, both the language understanding and generation processes need to handle equivariance properties in data. For example, the narrative structure of an article can be reorganized, while still presenting the same content. In constrained NLG tasks with structural priors (e.g. structured data-to-text generation), the structure of the prior can also be modified while presenting semantically equivalent content. However, existing sequential modeling of languages cause downstream information extraction and NLG systems to be brittle to content-neutral transformations of input data. Our pilot study realizes equivariance learning by incorporating structured masking and transformation-invariant position encoding mechanisms in pre-trained Transformer models for data-to-text [44] and scene-to-text [49] generation tasks. Following this direction, I will investigate principled approaches to address three research questions: (i) How to capture implicit but semantically equivalent structures (e.g., narrative structures) of natural language text, and accordingly generate equivariant language representations; (ii) How to disentangle semantic and syntactic representation in large language models; (iii) How to composite information from multiple components of text (e.g., sentences, paragraphs, or documents) while ensuring the equivariance to positional and frequential perturbations.

Cross-domain and Interdisciplinary Research. I always believe that a useful technology should address problems in several related research areas rather than a single one. Therefore, beyond core NLP tasks, my research has also contributed to computational biology [4, 61, 22], medical informatics [58, 57, 20], software engineering [23, 24, 60], geo-intelligence [26] and social media analysis [40, 1]. Particularly, I have been interested in AI technologies for common good that could contribute to fairness [62], healthcare [58, 57] and education [31]. Given the previous success in transferring technologies to different areas, I am enthusiastic about developing open-source libraries and software, and facilitating collaborations with people outside my areas. I am excited about any opportunities to apply my expertise in NLP and representation learning to solve important problems in other areas and disciplines.

References

[1] CHEN, H., SULTAN, S. F., TIAN, Y., CHEN, M., AND SKIENA, S. Fast and accurate network embed-

dings via very sparse random projection. In *CIKM* (2019).

[2] CHEN, M., CHANG, K.-W., AND ROTH, D. Recent advances in transferable representation learning. In *AAAI Tutorials* (2020).

[3] CHEN, M., HUANG, L., LI, M., ZHOU, B., JI, H., AND ROTH, D. New frontiers of information extraction. In *NAACL Tutorials* (2022).

[4] CHEN, M., JU, C., ZHOU, G., CHEN, X., ZHANG, T., CHANG, K.-W., ZANIOLO, C., AND WANG, W. Multifaceted protein-protein interaction prediction based on siamese residual rcnn. *Bioinformatics 35*, 14 (07 2019), i305–i314 (Procs of ISMB 2019).

[5] CHEN, M., AND QUIRK, C. Embedding edge-attributed relational hierarchies. In *SIGIR* (2019).

[6] CHEN, M., SHI, W., ZHOU, B., AND ROTH, D. Cross-lingual entity alignment with incidental supervision. In *EACL* (2020).

[7] CHEN, M., SHI, W., ZHOU, P., AND CHANG, K.-W. Retrofitting contextualized word embeddings with paraphrases. In *EMNLP* (2019).

[8] CHEN, M., TIAN, Y., CHANG, K.-W., SKIENA, S., AND ZANIOLO, C. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment. In *IJCAI* (2018).

[9] CHEN, M., TIAN, Y., CHEN, H., CHANG, K.-W., SKIENA, S., AND ZANIOLO, C. Learning to represent bilingual dictionaries. In *CoNLL* (2019).

[10] CHEN, M., TIAN, Y., ET AL. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *IJCAI* (2017).

[11] CHEN, M., TIAN, Y., ET AL. On2vec: Embedding-based relation prediction for ontology population. In *SDM* (2018).

[12] CHEN, M., ZHANG, H., NING, Q., LI, M., JI, H., MCKEOWN, K., AND ROTH, D. Event-centric natural language processing. In *ACL Tutorials* (2021).

[13] CHEN, M., ZHANG, H., NING, Q., LI, M., JI, H., AND ROTH, D. Event-centric natural language understanding. In *AAAI Tutorials* (2021).

[14] CHEN, M., ZHANG, H., WANG, H., AND ROTH, D. What are you trying to do? semantic typing of event processes. In *CoNLL* (2020), **Best Paper Nomination**.

[15] CHEN, M., ZHOU, T., ET AL. Multi-graph affinity embeddings for multilingual knowledge graphs. In *AKBC* (2017).

[16] CHEN, X., BORATKO, M., CHEN, M., DASGUPTA, S. S., LI, X. L., AND MCCALLUM, A. Probabilistic box embeddings for uncertain knowledge graph reasoning. In *NAACL* (2021).

[17] CHEN, X., CHEN, M., FAN, C., UPPUNDA, A., AND ZANIOLO, C. Cross-lingual knowledge graph completion via ensemble knowledge transfer. In *EMNLP* (2020).

[18] CHEN, X., CHEN, M., SHI, W., SUN, Y., AND ZANIOLO, C. Embedding uncertain knowledge graph. In *AAAI* (2019).

[19] HAO, J., CHEN, M., YU, W., SUN, Y., AND WANG, W. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts. In *KDD* (2019).

[20] HAO, J., JU, C. J.-T., CHEN, M., SUN, Y., ZANIOLO, C., AND WANG, W. Bio-joie: Joint representation learning of biological knowledge bases. In *ACM BCB* (2020), **SIGBio Best Student Paper**.

[21] HUANG, J. Y., LI, B., XU, J., AND CHEN, M. Unified semantic typing with meaningful label inference. In *NAACL* (2022).

[22] JIANG, J.-Y., JU, C. J.-T., HAO, J., CHEN, M., AND WANG, W. Jedi: circular rna prediction based on junction encoders and deep interaction among splice sites. *Bioinformatics 37*, Supplement_1 (2021), i289–i298 (Procs of ISMB 2019).

[23] KOU, B., DI, Y., CHEN, M., AND ZHANG, T. Sosum: A dataset of stack overflow post summaries. In *MSR* (2022).

[24] KOU, B., DI, Y., CHEN, M., AND ZHANG, T. Automated summarization of stack overflow posts. In *ICSE* (2023), in submission.

[25] LI, B., YIN, W., AND CHEN, M. Ultra-fine entity typing with indirect supervision from natural language inference. *TACL 10* (2022), 607–622.

[26] LI, Z., KIM, J., CHIANG, Y.-Y., AND CHEN, M. Spabert: Pretrained language models on geographic data for geo-entity representation. In *EMNLP* (2022), in submission.

[27] LIU, F., CHEN, M., ROTH, D., AND COLLIER, N. Visual pivoting for (unsupervised) entity alignment. In *AAAI* (2021).

[28] LIU, J., SUN, Z., HOOI, B., WANG, Y., LIU, D., YANG, B., XIAO, X., AND CHEN, M. Dangling-aware entity alignment with mixed high-order proximities. In *NAACL - Findings* (2022).

[29] LU, K., HSU, I.-H., MA, M. D., ZHOU, W., AND CHEN, M. Summarization as indirect supervision for relation extraction. In *EMNLP* (2022), in submission.

[30] MA, M. D., CHEN, M., WU, T.-L., AND PENG, N. Hyperexpan: Taxonomy expansion with hyperbolic representation learning. In *EMNLP - Findings* (2021).

[31] MENG, C., CHEN, M., MAO, J., AND NEVILLE, J. Readnet: A hierarchical transformer framework for readability analysis. In *ECIR* (2020).

[32] PAULHEIM, H. How much is a triple? estimating the cost of knowledge graph creation. In *ISWC* (2018).

[33] QASEMI, E., KHANNA, P., NING, Q., AND CHEN, M. Pinks: Preconditioned commonsense inference with minimal supervision. In *AACL* (2022), in submission.

[34] QASEMI, E., LIU, F., CHEN, M., AND SZEKELY, P. Paco: Preconditions attributed to commonsense knowledge. In *EMNLP* (2022), in submission.

[35] SUN, Z., CHEN, M., AND HU, W. Knowing the no-match: Entity alignment with dangling cases. In *ACL* (2021).

[36] SUN, Z., CHEN, M., HU, W., WANG, C., DAI, J., AND ZHANG, W. Knowledge association with hyperbolic knowledge graph embeddings. In *EMNLP* (2020).

[37] SUN, Z., HUANG, J., HU, W., CHEN, M., AND QU, Y. Transedge: Translating relation-contextualized embeddings for knowledge graphs. In *ISWC* (2019).

[38] SUN, Z., WANG, C., HU, W., CHEN, M., DAI, J., ZHANG, W., AND QU, Y. Knowledge graph alignment network with gated multi-hop neighborhood aggregation. In *AAAI* (2020).

[39] SUN, Z., ZHANG, Q., HU, W., WANG, C., CHEN, M., AKRAMI, F., AND LI, C. A benchmarking study of embedding-based entity alignment for knowledge graphs. *PVLDB 13* (2020).

[40] TIAN, Y., CHEN, H., CHEN, M., PEROZZI, B., AND SKIENA, S. Social relation inference via label propagation. In *ECIR* (2019).

[41] WANG, F., HUANG, K.-H., CHANG, K.-W., AND CHEN, M. Self-augmentation improves zero-shot cross-lingual transferability. In *EMNLP* (2022), in submission.

[42] WANG, F., SONG, K., ZHANG, H., CHO, S., YAO, W., CHEN, M., AND YU, D. Salience allocation as guidance for abstractive summarization. In *EMNLP* (2022), in submission.

[43] WANG, F., SUN, K., PUJARA, J., SZEKELY, P. A., AND CHEN, M. Table-based fact verification with salience-aware learning. In *EMNLP - Findings* (2021).

[44] WANG, F., XU, Z., SZEKELY, P., AND CHEN, M. Robust (controlled) table-to-text generation with structure-aware equivariance learning. In *NAACL* (2022).

[45] WANG, H., CHEN, M., ZHANG, H., AND ROTH, D. Joint constrained learning for event-event relation extraction. In *EMNLP* (2020).

[46] WANG, H., ZHANG, H., CHEN, M., AND ROTH, D. Learning constraints and descriptive segmentation for subevent detection. In *EMNLP* (2021).

[47] WANG, H., ZHANG, H., DENG, Y., ROTH, D., AND CHEN, M. Extracting or guessing? improving faithfulness of event temporal relation extraction. In *EMNLP* (2022), in submission.

[48] WANG, P., ILIEVSKI, F., CHEN, M., AND REN, X. Do language models perform generalizable commonsense inference? In *ACL - Findings* (2021).

[49] Wang, P., Zamora, J., Liu, J., Liu, F., Chen, M., and Ren, X. Contextualized scene imagination for generative commonsense reasoning. In *ICLR* (2022).

[50] Wang, Y., Chen, M., Zhou, W., Cai, Y., Liang, Y., Liu, D., Yang, B., Liu, J., and Hooi, B. Should we rely on entity mentions for relation extraction? debiasing relation extraction with counterfactual analysis. In *NAACL* (2022).

[51] Xiong, W., Li, X., Iyer, S., Du, J., Lewis, P., Wang, W. Y., Mehdad, Y., Yih, S., Riedel, S., Kiela, D., et al. Answering complex open-domain questions with multi-hop dense retrieval. In *ICLR* (2021).

[52] Xu, N., Wang, F., Li, B., Dong, M., and Chen, M. Does your model classify entities reasonably? diagnosing and mitigating spurious correlations in entity typing. In *EMNLP* (2022), in submission.

[53] Yang, X., Huang, J. Y., Zhou, W., and Chen, M. Parameter-efficient tuning with special token adaptation. In *EMNLP* (2022), in submission.

[54] Zhang, H., Chen, M., Wang, H., Song, Y., and Roth, D. Analogous process structure induction for sub-event sequence prediction. In *EMNLP* (2020).

[55] Zhang, H., Wang, Y., Deng, Y., Wang, H., Chen, M., and Roth, D. Are all steps equally important? benchmarking essentiality detection of events. In *ACL ARR* (2022), in submission.

[56] Zhang, Q., Sun, Z., Chen, M., Guo, L., and Qu, Y. Multi-view knowledge graph embedding for entity alignment. In *IJCAI* (2019).

[57] Zhang, T., Chen, M., and Bui, A. Diagnostic prediction with sequence-of-sets representation learning for clinical event. In *AIME* (2020).

[58] Zhang, T., Chen, M., and Bui, A. Adadiag: Adversarial domain adaptation of diagnostic prediction with clinical event sequences. *Journal of Biomedical Informatics (JBI)* (2022).

[59] Zhang, X., Chen, M., and May, J. Salience-aware event chain modeling for narrative understanding. In *EMNLP* (2021).

[60] Zhao, Z., Ibrahim, M. Y., Chen, M., and Zhang, T. Extracting version compatibility knowledge from stack overflow via indirect supervision. In *ICSE* (2023), in submission.

[61] Zhou, G., Chen, M., Ju, C., Wang, Z., Jiang, J.-Y., and Wang, W. Mutation effect estimation on proteinprotein interactions using deep contextualized representation learning. *NAR Genom. Bioinform.* (2020).

[62] Zhou, P., Shi, W., Zhao, J., Huang, K.-H., Chen, M., Cotterell, R., and Chang, K.-W. Examining gender bias in languages with grammatical gender. In *EMNLP* (2019).

[63] Zhou, W., and Chen, M. Learning from noisy labels for entity-centric information extraction. In *EMNLP* (2021).

[64] Zhou, W., Liu, F., and Chen, M. Contrastive out-of-distribution detection for pretrained transformers. In *EMNLP* (2021).

[65] Zhou, W., Liu, F., Zhang, H., and Chen, M. Sharpness-aware minimization with dynamic reweighting. In *EMNLP* (2022), in submission.

[66] Zhou, W., Ning, Q., Elfardy, H., Small, K., and Chen, M. Answer consolidation: Formulation and benchmarking. In *NAACL* (2022).

[67] Zhu, C., Chen, M., Fan, C., Cheng, G., and Zhang, Y. Learning from history: Modeling temporal knowledge graphs with sequential copy-generation networks. In *AAAI* (2021).