WSDM 2023 Tutorial

Concluding Remarks: Knowledge-Augmented Methods for NLP

Meng Jiang
University of Notre Dame

Three Slides: Integrating Knowledge in NLP

- **Augment Knowledge for What? (a) Natural Language Understanding**
  - Text classification: Sentiment analysis, Fact verification
  - Information extraction: NER, Entity linking, Slot filling, Relation prediction
  - Question answering: Open-domain QA, Commonsense QA, Knowledge-base QA
- **Obtain Knowledge from Where?**
  - Wiki-based: Wikipedia, WikiData, Wiktionary
  - General domain: Freebase, DBpedia, YAGO
  - Specific domains: UMLS, ArnetMiner, DBLP
  - Commonsense: OMCS, ConceptNet
- **Use Knowledge via How?**
  - Entity linking based methods
    - ERNIE (ACL'19), KEAR (IJCAI'22), EaE (EMNLP'20), FILM (ACL'21), K-BERT (AAAI'20)
  - Retrieval based methods
    - DPR (EMNLP'20), REALM (ICML'20), REINA (ACL'22), RETRO ('21), WebGPT ('21)

Three Slides: Integrating Knowledge in NLP

- **Augment Knowledge for What? (b) Natural Language Generation**
  - Question answering: Question generation and Answer generation
  - Dialog systems: Response generation
  - Reasoning: Explanation generation
  - Machine translation; Summarization; Paraphrasing
- **Obtain Knowledge from Where?**
  - Structured knowledge (Knowledge graph): WikiData, Freebase, DBPedia, YAGO, ConceptNet
  - Unstructured knowledge (Grounded document): Wikipedia, Wiktionary, ArnetMiner, OMCS
- **Use Knowledge via How?**
  - Knowledge graph based methods
    - GRF (EMNLP'20), CCM (IJCAI'18), MoKGE (ACL'22)
  - Grounded document based methods
    - RAG (NeurIPS'20), RE-T5 (ACL'21), CMR (ACL'19)

Three Slides: Integrating Knowledge in NLP

- **Augment Knowledge for What? (c) Commonsense Reasoning**
  - "commonsense reasoning is a human-like ability to make presumptions about the type and essence of ordinary situations humans encounter every day."
  - Human-level AI
- **Obtain Knowledge from Where?**
  - Commonsense: OMCS, ConceptNet
- **Use Knowledge via How?**
  - KagNet (EMNLP'19), MHGRN (EMNLP'20), QA-GNN (NAACL'21)
  - GreaseLM (ICLR'22), GSC (ICLR'22)
  - CommonGen (EMNLP'20), KFCNet (EMNLP'21), KG-BART (AAAI'21), I&V (ICLR'22)
  - DrFact (NAACL'21): Concept-Fact Hypergraph, Dense fact embeddings

Future Directions

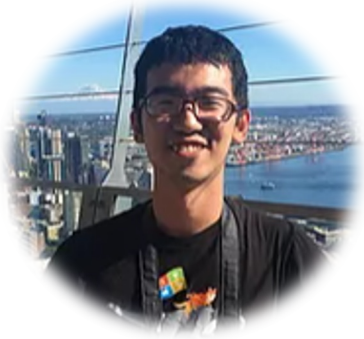
- Augment Knowledge for What?
  - **Desired properties: Accuracy, Diversity, Interpretability, etc.**
  - **Specialized domains: Technical support, Online education, Emotional support, Scientific discovery, etc.**
- Obtain Knowledge from Where?
  - **Heterogeneity: Multiple types of knowledge source data**
  - **Interconnectivity: Multiple knowledge fragments that are complementary or interconnected**
  - **Veracity: Multiple levels of reliability of information sources (Beyond verified information)** | **Scalability, etc.**
- Use Knowledge via How? **Your MAGIC!**

Contact

Chenguang Zhu

Principal Research Manager

Microsoft Cognitive Services Research



Yichong Xu

Senior Researcher

Microsoft Cognitive Services Research

Xiang Ren

Assistant Professor

Dept. of Computer Science

USC

Bill Yuchen Lin

Young Investigator

Allen Institute for AI

Meng Jiang

Assistant Professor

Dept. of Computer Science and Engineering

University of Notre Dame

Wenhao Yu

Ph.D. candidate

Dept. of Computer Science and Engineering

University of Notre Dame

Acknowledgements

IIS-1849816, IIS-2119531, IIS-2142827, IIS-2146761, IIS-2137396, CCF-1901059