CS224V: Conversational Virtual Assistants with Deep Learning

Reading List
Fall 2023

1 Large Language Models (LLMs)
   Introduction
   1. Attention (Vaswani et al., 2017)
   2. GPT-3 (Brown et al., 2020)
   3. Instruct-GPT (Ouyang et al., 2022)
   4. LLAMA (Touvron et al., 2023a)
   5. Alpaca (Taori et al., 2023)
   6. Alpaca with Self-Instruct (Wang et al., 2023c)
   7. LLAMA-2 (Touvron et al., 2023b)
   8. Chain-of-thought (Wei et al., 2023)
   9. Self-consistency (Wang et al., 2023b)

2 Grounding LLMs on Free Text
   2.1 Neural Information Retrieval Models
      Non-neural algorithms: TF-IDF and BM-25.
      Popular neural retrieval systems:
      1. ColBERT (Khattab and Zaharia, 2020)
      2. Condenser (Gao and Callan, 2021)
      3. CoCondenser (Gao and Callan, 2022)
      4. CoCo-DR (Yu et al., 2022)
   2.2 Retrieval + Generation
      1. Citation generation (Gao et al., 2023)
      2. Active retrieval augmented generation (Jiang et al., 2023)
      3. WikiChat (Semnani et al., 2023)
   2.3 Evaluation
      1. Evaluating Verifiability in Generative Search Engines (Liu et al., 2023a)
      2. Generating Benchmarks for Factuality Evaluation (Muhlgay et al., 2023)

3 Grounding LLMs on Databases, Knowledge Graphs, and heterogeneous sources
   1. Schema2QA (Xu et al., 2020)
   2. Grail QA (Gu et al., 2021)
   3. BIRD: Text-to-SQL benchmark for LLMs (Li et al., 2023a)
   4. WikiData semantic parser (Xu et al., 2023)
   5. Compmix: a heterogeneous data set with WikiData and Wikipedia (Christmann et al., 2023)
   6. Named Entity Disambiguation (NED): Re-FinED (Ayoola et al., 2022)

4 Multi-Modal Applications
   1. React: Describing the UI. (React)
   2. ReactGenie Framework for Multimodal Applications (Yang et al., 2023)

5 Task-Oriented Dialogue Agents
   1. MultiWOZ (Budzianowski et al., 2018)
   2. Dialogue Agent Architecture (Campagna et al., 2022)
   3. RiSAWOZ dataset (Chinese) (Quan et al., 2020)
   4. X-RiSAWOZ multilingual dataset (Moradshahi et al., 2023)

6 Social Agents
   1. Persuasion for Good: Towards a Personalized Persuasive Dialogue System for Social Good (Wang et al., 2019)
   2. Controllable mixed-initiative dialogue generation through prompting (Chen et al., 2023)

4. Cardinal Chirpy (Chi et al., 2021)

5. Blenderbot (Shuster et al., 2022)

7 Robotic Automation

1. Russ: Grounding Open-Domain Instructions to Automate Web Support Tasks (Xu et al., 2021)

2. DIY assistant: a multi-modal end-user programmable virtual assistant (Fischer et al., 2021)

8 Grounding Agents on APIs and DSLs

8.1 Tools and APIs

1. ToolFormer (Schick et al., 2023)

2. ART: Multi-step tool use (Paranjape et al., 2023)

3. Gorilla LM (Patil et al., 2023)

4. ToolAlpaca (Tang et al., 2023)

8.2 Domain-Specific Languages (DSL)

1. Event-driven execution (Campagna et al., 2017)

2. Access control using satisfiability modulo theory (Campagna et al., 2018)

9 Large Language Models

9.1 Distillation of LLMs

1. Chain-of-Thought distillation (Li et al., 2023b)

2. SCOTT: Self-consistent Chain-of-Thought distillation (Wang et al., 2023a)

3. Symbolic Commonsense Knowledge Distillation (West et al., 2022)

4. Knowledge Distillation of Large Language Models (Qiu et al., 2023)

5. Evaluating Open-Domain Question Answering in the Era of Large Language Models (Kamaloo et al., 2023)

6. Self-Refine (Madaan et al., 2023)

9.2 Evaluation of LLMs

1. HELM (Liang et al., 2022),

2. Repairing the Cracked Foundation: A Survey of Obstacles in Evaluation Practices for Generated Text (Gehrmann et al., 2022)

3. Judging LLM-as-a-judge with MT-Bench and Chatbot Arena (Zheng et al., 2023)

4. G-Eval: NLG Evaluation using GPT-4 with Better Human Alignment (Liu et al., 2023b)

10 Curation of Common Sense Knowledge

1. (Comet-) atomic 2020: On symbolic and neural commonsense knowledge graphs (Hwang et al., 2021)

2. Commonsense Knowledge Transfer for Pre-trained Language Models (Zhou et al., 2023)

References


Tianyu Gao, Howard Yen, Jiatong Yu, and Danqi Chen. 2023. Enabling large language models to generate text with citations.


Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. 2023. Knowledge distillation of large language models.


Jinyang Li, Binyuan Hui, Ge Qu, Binhua Li, Jiaxi Yang, Bowen Li, Bailin Wang, Bowen Qin, Rongyu Cao, Ruiyiing Geng, Nan Hua, Xuanhe Zhou, Chenhao Ma, Guoliang Li, Kevin C. C. Chang, Fei Huang, Reynold Cheng, and Yongbin Li. 2023a. Can ilm already serve as a database interface? a big bench for large-scale database grounded text-to-sqls.


Yang Liu, Dan Iter, Yichong Xu, Shuohang Wang, Ruochen Xu, and Chenguang Zhu. 2023b. G-eval: NLig evaluation using gpt-4 with better human alignment.


Dor Muhlgay, Ori Ram, Inbal Magar, Yoav Levine, Nir Ratner, Yonatan Belinkov, Omri Abend, Kevin Leyton-Brown, Amnon Shashua, and Yoav Shoham. 2023. Generating benchmarks for factuality evaluation of language models.


Qiaoyu Tang, Ziliang Deng, Hongyu Lin, Xianpei Han, Qiao Liang, and Le Sun. 2023. Toolalpaca: Generalized tool learning for language models with 3000 simulated cases.


Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajiwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillemin Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Sagar Hosseini, Rui Hou, Hakan


