

Pengshan Cai

Pursuing Positions in NLP @ Summer 2023

☎ 413-475-8552 | ✉ pengshancai@umass.edu

Educations

University of Massachusetts, Amherst

Sep. 2017 -

- PhD student in Computer Science

Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

Sep. 2014 - June.2017

- M.E. in Computer Science and Technology

Wuhan University, Wuhan, China

Sep. 2010 - June.2014

- B.E. in Computer Science and Technology

Publication

Pengshan Cai, Kaiqiang Song, Sangwoo Cho, Hongwei Wang, Xiaoyang Wang, Hong Yu, Fei Liu and Dong Yu *Generating User-Engaging News Headlines* ACL 2023

Pengshan Cai, Fei Liu, Adarsha Bajracharya, Weisong Liu, Dan Berlowitz, Joe Sills, Alok Kapoor, Richeek Pradhan, David Levy, Hong Yu *Generation of Patient After-Visit Summaries to Support Physicians* COLING 2022

Pengshan Cai, Hui Wan, Fei Liu, Mo Yu, Hong Yu, Sachindra Joshi *Learning as Conversation: Dialogue Systems Reinforced for Information Acquisition* NAACL 2022

Michael Glass, Gaetano Rossiello, Md Faisal Mahbub Chowdhury, Ankita Naik, Pengshan Cai, Alfio Gliozzo *Re2G: Retrieve, Rerank, Generate* NAACL 2022

Zhichao Yang*, Pengshan Cai*, Yansong Feng, Fei Li, Weijiang Feng, Elena Chiu, Hong Yu. *Generating Classical Chinese Poems from Vernacular Chinese* (equal first author) EMNLP 2019

Pengshan Cai, Yansong Feng, Yantao Jia, Yuanzhuo Wang, Xiaolong Jin, Xueqi Cheng. *Coarse to Fine: Diffusing Categories in Wikipedia* WWW 2018

Denghui Zhang, Manling Li, Pengshan Cai, Yantao Jia, Yuanzhuo Wang. *Path-Based Attention Neural Model for Fine-Grained Entity Typing* AAAI 2018

Pengshan Cai, Mo Yu, Fei Liu, Hong Yu *Generating Coherent Narratives with Subtopic Planning to Answer How-to Questions* EMNLP Workshop on Generation, Evaluation Metrics (GEM) 2022

Pengshan Cai, Wei Li, Yansong Feng, Yuanzhuo Wang, Yantao Jia *Learning Knowledge Representation Across Knowledge Graphs* AAAI Workshop on Knowledge-Based Techniques for Problem Solving and Reasoning 2017

Projects

Personalized Headline Generation for News Recommendation

May. 2022 -

- Problem: Readers decide to read a news passage by reading its headline, which may not catch the users' interests.
- Exploring personalized summarization models: Generate news headlines which not only catch user interests, but also loyal to the original passage.
- The related paper was accepted by ACL 2023 (Lead author).

No Need to Read, Just Talk

May. 2021 - Dec. 2021

- Problem: As a passive way to gain information, reading passages consumes vast intellectual efforts, and poses even more challenges to people with attention disorders (e.g. ADHD)
- Propose a novel task: Letting a dialogue agent read a passage and then have a conversation with the user. Through the conversation, the user is expected to gain the information in the passage.
- Propose a reinforcement learning based fine-tuning approach, which attempt to improve the chat bot on coverage and coherence, the experiment results demonstrate the fine-tuned chat bot could carry out knowledgeable and attentive conversations
- The related paper was accepted by NAACL 2022 (Lead author).

Supporting Clinicians in the Generation of Patient After-Visit Summaries

Jan. 2022 - May. 2022

- Problem: While automatic summarization models could be applied to help generate after-visit summaries for patients, the generated summaries usually contain many errors, and could be directly given to patients without physicians' revision
- We propose a novel approach to automatically identify two types of common errors in the generated after-visit summaries: Missing content errors and factual errors, the proposed system help physicians identify and correct errors.
- With the help of the proposed model, physicians are able to write after-visit-summaries for patients with higher efficiency.
- The related paper was accepted by COLING 2022 (Lead author).

Interactive Dialogue Agent for Patient Education

Feb. 2022 -

- Problem: Patient may have difficulty understanding their clinical records, which leads to poor self-management.
- Propose a novel task: Letting patients have conversations with a dialogue agent. Through the conversation, the dialogue agent asks questions based on the patients' clinical records, and correct patients once it detects wrong answers.
- Submitted to TACL (Equal lead author).

Generating Classical Chinese Poems from Vernacular Chinese

Feb. 2019 - May. 2019

- Problem: In most previous poem generation researches, the semantic of the generated poems could not be well controlled by users.
- Exploring reinforcement learning based models to allow users better control the semantic of generated poems, as well as ways to alleviate over-translation and under-translation in poem generation.
- The related paper was accepted by EMNLP 2019 (Equal Lead author).

Learning to Diffuse Categories in Wikipedia

Feb. 2016 - Jul. 2016

- Problem: A large number of categories in Wikipedia are coarse-grained. Users expect a finer-grained categorization system which could provide them with more information.
- Based on defining attribute theory in cognitive psychology, proposed a model to automatically find the defining-attributes of a coarse-grained category in Wikipedia, thus generate finer-grained subcategories from the coarse-grained category.
- The related paper was accepted by WWW 2017 (Lead author).

Experience

Tencent AI Lab

RESEARCH INTERN

Bellevue, WA

May 2022 - Aug. 2022

- Researches concerning personalized summarization

IBM Research AI

RESEARCH INTERN

Yorktown Heights, NY

May 2021 - Aug. 2021

- Researches concerning dialogue system

Bytedance Inc.

RESEARCH INTERN

Mountain View, CA

May 2020 - Aug. 2020

- Researches concerning automatic lyric generation

Siemens Healthineers

RESEARCH INTERN

Princeton, NJ

May 2019 - Aug. 2019

- Building automatic Q&A system for customer service

Baidu Inc.

RESEARCH INTERN

Beijing, China

May 2018 - Aug. 2018

- Working in the medical NLP group on few-shot learning for relation classification

Institute of Computer Science and Technology, Peking University

RESEARCH ASSISTANT

Beijing, China

Dec. 2015 - Jun. 2017

- Researches concerning knowledge graph