Shahla Farzana

sfarza3@uic.edu
https://www.linkedin.com/in/shahlafarzana/ 864-6339254
https://github.com/treena908

Research Summary

My primary research interests are in: (1) developing and evaluating multi-modal and multi-lingual (speech and text) computational approaches for low-resource tasks in cognitive and behavioral health, and (2) dialogue systems and their applications. The overarching goal of my research is to leverage the power of data and recent advancements in NLP to build automated, accessible systems for cognitive and psychological well-being as well as for educational and business platforms.

EDUCATION

Ph.D. Candidate, Computer Science *Spring, 2019 - Present*
University of Illinois at Chicago (UIC)
Supervisor : Dr. Natalie Parde
Exp. Graduation Date: March, 2024

B.Sc., Computer Science & Engineering *2010 - 2015*
Bangladesh University of Engineering & Technology (BUET)

TECHNICAL SKILLS

Languages : Proficient: JAVA, Python, R **Advanced:** C, C++, PHP
Databases : Oracle, MySQL
Tools/Frameworks : PyTorch, Tensorflow, Gensim, Dialogflow, INCEPTION (Annotation Tool), Adobe XD, Draw.io
Language Models: BERT, GPT-2, T-5

PROJECTS

Modeling Buyers Behaviour for Key Entity Identification in User Query
| *Scala, PySpark, PyTorch* | *Su'23*

- Created dataset from users query reformulation log where the rewritten query is hypothesised to preserve the key entities of the original query after dropping more than 50% of the original tokens. We experimented with finetuning Bert models (eBert, microBert) as a sequence classification model to rewrite queries given the original ones. Evaluating the system against the online inventory with recall/relevance model, we found performance improvement (recall and relevance) when applying the key entities predicted by the model compared to the original query. Moreover, we empirically find that applying key entity constraints on embedding based retrieved items improves the overall relevance of the system.

Neural Query Rewriting | *PyTorch* | *Su'22*

- Developed and experimented with Knowledge Graph (KG) enhanced neural query rewriting framework to solve null and low query issue of e-commerce search engine and achieved 2% increase in F1 score over the baseline model.

Dementia Detection
| *scikit-learn, ADAPT, Tensorflow, Keras, PyTorch* | *2019-2022*

- We experiment with domain adaptation (DA) techniques on heterogeneous spoken language data to evaluate generalizability across diverse datasets for a common task: dementia detection. We empirically find that the feature-augmented DA method achieves a 22% increase in accuracy adapting from a conversational to task-specific dataset compared to a jointly trained baseline. [Link]
- Annotated conversational data with Dialog-Act (DA) tags from two different cognitive tasks of *DementiaBank* dataset, proposed task-agnostic Dementia detection framework leveraging the annotations, and achieved competitive performance relative to the models trained on a specific cognitive task. [Link]

- An automated system to predict MMSE scores reflecting individuals' cognitive health status, based on their free speech samples from conversational interviews. The models experiment with traditional machine learning algorithms, pre-trained language model, and the best performing model achieved 16.5% decrease in RMSE score from the linguistic benchmark of the baseline paper in **ADReSS Challenge2020**. [Link]

Mini-Alexa | *SQLite, Python, Stanford CoreNLP Parser* | *2019*

- A rule-based textual question answering (QA) system capable of answering short questions on a specific domain. [Link]

Virtual-Interviewer Chatbot | *Dialogflow, Node.js* | *2019*

- A dialog agent designed to facilitate cognitive health screening task (*Picture Description*) with patients at risk for age-related dementia. [Link]

PUBLICATIONS

Conference Proceedings

- Shahla Farzana and Natalie Parde. Towards Domain-Agnostic and Domain-Adaptive Dementia Detection from Spoken Language. In the Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023). Toronto, Canada, July 9-14, 2023.
- Shahla Farzana and Natalie Parde. Are Interaction Patterns Helpful for Task-Agnostic Dementia Detection? An Empirical Exploration To appear in the Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2022). Edinburgh, Scotland, September 7-9, 2022.
- Shahla Farzana and Natalie Parde. Exploring MMSE Score Prediction Using Verbal and Non-Verbal Cues. In the Proceedings of the 21st Conference of the International Speech Communication Association (INTERSPEECH 2020). Shanghai, China, October 25-29, 2020.
- Shahla Farzana, Mina Valizadeh, and Natalie Parde. Modeling Dialogue in Conversational Cognitive Health Screening Interviews. In the Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020). Marseilles, France, May 11-16, 2020.
- Shahla Farzana, Khaleda Akther Papry, Ashikur Rahman, and Raqeebir Rab. Maximally Pair-wise Disjoint Set Covers for Directional Sensors in Visual Sensor Networks. In Proceedings of IEEE/IFIP Wireless Days (WD16), March 23-25, Touolouse, France, 2016. Acceptance Rate: 33.6%.

Workshop Proceedings

- Shahla Farzana, Qunzhi Zhou, Petar Ristoski. "Knowledge Graph-Enhanced Neural Query Rewriting". To appear in the Proceedings of the 2nd International Workshop on Interactive and Scalable Information Retrieval Methods for e-Commerce (ISIR-eCom 2023).
- Shahla Farzana, Ashwin Deshpande, and Natalie Parde. "How You Say It Matters: Measuring the Impact of Verbal Disfluency Tags on Automated Dementia Detection". In the Proceedings of the 21st Workshop on Biomedical Language Processing (BioNLP 2022). Dublin, Ireland, May 26, 2022.

PRESENTATIONS

• **eBay Intern Showcase**, Modeling Buyers Behaviour for Key Entity Identification in User Query *Su'23*
• **WWW'23 Conference**, Neural Query Rewriting. *Su'23*
• **ACL Conference**, Towards Domain-Agnostic and Domain-Adaptive Dementia Detection from Spoken Language. *Su'23*
• **SIGdial Conference** (virtual), Are Interaction Patterns Helpful for Task-Agnostic Dementia Detection? An Empirical Exploration *Su'22*

• **INTERSPEECH Conference** (Virtual), Exploring MMSE Score Prediction Using Verbal and Non-Verbal Cues. *Sp'20*

AWARDS & Achievements

• **Finalist**, 2023 UIC 3MT Thesis Competition (one of the 13 finalists among 50 applicants)
• **Scholarships:** ACL Student Volunteer grant 2023, ISCA student grant 2020 (Interspeech conference), GHC 2020, CRA-W Grad Cohort Workshop 2020
• UIC Graduate Student Council (GSC) travel award 2023, SC16 student volunteer, GHC India (GHCI) 2014
• **Dean's List Award**, for excellent results at BUET *2013*

SERVICES

• **Student Volunteer**, ACL *2023*
• **Research Panelist**, Women in Computer Science (WiCS) club of UIC *2022*
• **Reviewer**, CIKM, Computer Speech and Language Journal *SU'22, Sp'21*
• **Graduate Student Council Representative**, UIC, USA *F'19 - Sp'20*

REFERENCE

• Barbara Di Eugenio,Professor, Director of Graduate Studies, Department of Computer Science, email: bdieugen@uic.edu

• Natalie Parde, Assistant Professor, Department of Computer Science, University of Illinois Chicago, email: parde@uic.edu

• Erin Sundermann, Associate Adjunct Professor, Psychiatry, University of California San Diego, email: esundermann@ucsd.edu