School of Computing, Engineering, and the Built Environment
Edinburgh Napier University

PHD STUDENT PROJECT

Application instructions:
Detailed instructions are available at :
https://www.napier.ac.uk/research-and-innovation/doctoral-college/how-to-apply

Prospective candidates are encouraged to contact the Director of Studies (see details below) to discuss the project and their suitability for it.

Project details

Supervisory Team:

- DIRECTOR OF STUDY: Dr Md Zia Ullah (Email: m.ullah@napier.ac.uk)
- 2ND SUPERVISOR: Dr Dimitra Gkatzia

Subject Group: Computer Science

Research Areas: Computer Science, Artificial Intelligence, Data Science, Machine Learning

Project Title: Query Performance Prediction Combining Traditional and Neural Information Retrieval Models

Project description:

In the field of information retrieval (IR), query performance prediction (QPP) aims to predict the effectiveness of a system for a given search query without resorting to relevance judgments. QPP has various advantages, such as informing an IR system whether a given search query would be effective or ineffective. Based on the effectiveness of the query, the system can either apply a query reformulation or an adaptive retrieval configuration [1,4] or engage in an interactive session with the user (i.e., conversational search [5]) to understand user search intent and provide a better search experience.

Accurate estimation of query performance is a challenging problem. Existing QPPs are developed using the pre-retrieval features based on the collection statistics and/or the post-retrieval features based on the top-retrieved documents [2, 3] and evaluated considering the sparse retrieval models (e.g., BM25 or DFR) as reference system. Due to the advent of the language model (e.g., BERT [6]), neural

retrieval models have been proposed and shown to have better retrieval effectiveness [7]. However, there are very few QPP approaches developed using the neural IR model.

This Ph.D. project aims to develop query performance predictors for spare, dense retrieval models, and the combination of both approaches. Experiments could be conducted on standard TREC collections (e.g., MS MARCO and TREC Deep learning tracks) to demonstrate the effectiveness of the developed QPPs and compare them with the state-of-the-art approaches. The project will also aim to experiment and evaluate the applicability of QPPs in a conversational search context [5].

Prospective applicants are encouraged to contact the supervisor before submitting their applications. Applications should make it clear the project you are applying for and the name of the supervisor(s).

References:

1. Mothe J and Ullah MZ, Defining an Optimal Configuration Set for Selective Search Strategy – A Risk-Sensitive Approach, Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM 2021), 2021.
2. Déjean S, Ionescu RT, Mothe J, and Ullah MZ, Forward and backward feature selection for query performance prediction, The 35th ACM/SIGAPP Symposium on Applied Computing (SAC), 2020.
3. Chifu AG, Laporte L, Mothe J, and Ullah MZ, Query Performance Prediction Focused on Summarized Letor Features, The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2018.
4. Deveau R, Mothe J, Ullah MZ, Nie JY, Learning to Adaptively Rank Document Retrieval System Configurations, ACM Transactions of Information Systems (ACM TOIS), 41 pages, pp.3:1-3:41, Volume 37, Issue 1, 2019.
5. Dalton J, Xiong C, Kumar V, and Callan J, CAsT-19: A Dataset for Conversational Information Seeking, Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1985--1988, 2020.
6. Devlin, Jacob and Chang, Ming-Wei and Lee, Kenton and Toutanova, Kristina, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805, 2018.
7. Nogueira, Rodrigo and Cho, Kyunghyun, Passage Re-ranking with BERT, arxiv:1901.04085, 2019

Candidate characteristics

Education:

A first degree (at least a 2.1) ideally in Computer Science or Data Science

Subject knowledge:

Information retrieval (IR), Natural language processing (NLP), and Machine learning/Deep learning

Essential attributes:

- Experience of fundamental knowledge of IR, NLP, LLMs, and Machine learning/Deep learning
- Competent in Shell scripting, R, Python, Java, and PyTorch
- Knowledge of Information Retrieval and Deep Learning
- Good written and oral communication skills
- Strong motivation, with evidence of independent research skills relevant to the project
- Good time management

Desirable attributes:

- Experience of IR Tools, such as Terrier IR/Lemur Indri/Anserini/Lucene/Splade