

Online Topic Identification from Social media Text Streams

Project Description:

Text streams such as Twitter streaming API analysis becomes an important approach for discovering new topics during an on-going situation. This is an important task to learn about new rapid changes happening live to gain situational awareness from user-generated content during crises and emergencies. Automatic extraction of these new topics from huge amounts of documents in real-time is a challenging task. To address these issues, we aim to use the well-known Non-Negative Matrix Factorization (NMF) technique to discover new topics for sudden crises.

Duties/Activities:

The Intern will work on an existing baseline code for NMF and gets supervision from mentors in tweaking the code to match the designated problem. Then, the intern will test the developed code on real Twitter datasets.

Required Skills:

Python, R, or MATLAB, programming skills are required. Understand the basic machine learning algorithms is a plus.

Learning Opportunities:

The intern will have opportunity to work with scientist on promising research problems. Also, to understand how to formulate a physical problem into mathematical equations and programming language. In addition, learn how to use and develop non-negative matrix factorization (NMF) algorithm on real datasets.

Expected Team Size: 2-3

Mentors

Name: Mohamed Elshrif

Name: Muhammad Imran

email: melshrif@hbku.edu.qa

email: mimran@hbku.edu.qa