

Qualifier Exam for Debjyoti Paul

Problem 1

Social media data is enormous, but semi-private. List relevant social media data sources, and explain what is known about their sizes (in terms of storage space, and number of records), including both what is (probably) privately controlled by companies, and what is available for sufficiently-motivated and -resourced academic researchers.

Explain the state-of-the-art (with references to research papers) for scraping such semi-publicly accessible data sets, and what are the largest bottlenecks for such tasks.

Predict (using a machine learning / data mining techniques on the data above) what the total number of social media records available to researchers will be in 2022.

Problem 2

Deep reinforcement learning is a useful technique for realizing goal-oriented algorithms. It combines the idea of reinforcement algorithm with deep learning and has been shown to be very effective in many different application scenarios. Answer the following questions:

- 1) Please provide a detailed review of deep reinforcement learning.
- 2) How would you apply deep reinforcement learning for tuning the performance of a DB system? (i.e., like what OtterTune does, but using deep reinforcement learning). Please outline your problem formulation, overview of your approach, and a sketch analysis.

Problem 3