Seminar

Department of Computer Science

Dr. Koushik Sinha

Assistant Professor, Department of Computer Science

Designing a Human-Machine Hybrid Computing System for Unstructured Data

Date: Wednesday, November 2, 2016
Time: 2:00 – 3:00 p.m.
Location: EGRA309C, CS Conference Room

Abstract

Current machine algorithms for qualitative analysis of unstructured data (in the form of social media posts, audio and video, among others) do not perform well and show low accuracies due to the need for human-like intelligence. However, machines are easily scalable, fast and the quality of their output is predictable. On the other hand, though humans are much better than machine algorithms at image and video analysis, natural language text and speech processing, they are unfortunately unpredictable, slower and can be erroneous or even malicious as computing agents. Therefore, a task execution engine which can enable human-augmented cloud computing by intelligently orchestrating machine and human computing resources would be able to provide richer and superior analytics on unstructured data than either of the two types of computing agents in isolation.

We believe that a key aspect of enabling such analytics would be to provide guaranteed service level objectives, in terms of accuracy, time and budget. In this paper, we present a microtask scheduler with integrated *service level objectives* (SLO) management. With this goal, we have introduced two new decision parameters: H-M ratio and *microtask completion rate*. An early prototype has been built and validated through simulation with actual performance data collected from anonymous crowd workers on Amazon Mechanical Turk. Machine computation was done using Hewlett Packard's Autonomy IDOL while ground truth was established through the use of known, expert workers.

To the best of our knowledge, ours is the first work that attempts to simultaneously attempt to address the three SLO parameters of accuracy, budget and deadline for data-parallel microtasks.