

Summer Internship Requisitions

Summer Research Program Intern in Distributed Structured Prediction Algorithms, Requisition #1669

Position Description:

The recent explosion in social media data provides a rich source of information for activity, event and trend analysis. Effectively mining this massive data source requires new approaches to processing, storage and retrieval, and analytics. Several internship projects will individually focus on a specific aspect of information retrieval systems to collaboratively develop a working prototype designed to improve the efficiency of open source data analysis.

Structured prediction is a form of supervised machine learning that involves predicting structured outputs instead of real or discrete-valued outputs (as in traditional classification and regression problems). For example, in part-of-speech tagging, given an input sequence of words, the goal is to predict the corresponding sequence of part-of-speech tags (e.g., noun, verb, adjective). Training times can be lengthy for structured prediction algorithms. However, a key step can be parallelized (where predictions are made for each training example with the best model so far). This project will involve implementing distributed structured prediction algorithms to run on the Apache Spark cluster computing framework.

Requirements:

Candidates should be pursuing an undergraduate or graduate degree in computer science, mathematics, statistics, or a related field.

Desired Skills:

Candidates should have a background in machine learning, data mining and/or statistics. The candidates should also have a strong grasp of a high level programming language (e.g., Java, Scala, or Python).

Keywords:

Machine learning, data mining, cloud computing, big data

Summer Research Program Intern in Improved Social Media Mining, Requisition #1671

Position Description:

The recent explosion in social media data provides a rich source of information for activity, event and trend analysis. Effectively mining this massive data source requires new approaches to processing, storage and retrieval, and analytics. Several internship projects will individually focus on a specific aspect of information retrieval systems to collaboratively develop a working prototype designed to improve the efficiency of open source data analysis.

Mining datasets for similar items via pair wise comparison is intractable at scale for many computationally expensive algorithms. However, locality-sensitive hashing (LSH) can enable such bulk comparisons by hashing similar items to the same “buckets” for quick approximate nearest neighbor search. We seek a summer intern interested in examining and extending state of the art LSH approaches for exacting similar text meaning or image meaning from social media activities.

Requirements:

Candidates should be working toward a BS, MS, or PhD degree in Computer Science, Mathematics, or a related science or engineering field.

Desired Skills:

Candidates should be familiar with machine learning, data analysis and/or statistics, and be able to program effectively in a high level language (e.g. MATLAB, Python, C++, Java).

Keywords:

Machine learning, data mining, natural language processing, computer vision, and big data

*Summer Research Program Intern in Data Summarization, Requisition #1672***Position Description:**

The recent explosion in social media data provides a rich source of information for activity, event and trend analysis. Effectively mining this massive data source requires new approaches to processing, storage and retrieval, and analytics. Several internship projects will individually focus on a specific aspect of information retrieval systems to collaboratively develop a working prototype designed to improve the efficiency of open source data analysis.

Data summarization, a key concept in data mining, involves finding a compact description of a dataset. While these descriptions are often designed towards coherence, completeness, simplicity or some other measure of quality, they do not often account for memorability. By understanding how humans process and retain information we hope to design a more effective approach for summarizing large datasets. For example, consider the challenges in representing a collection of images using a fixed-length subsample. We expect that if a subsample contains exemplars that are more memorable, the resulting summarization is more effective. This project will combine methods from computer science, psychology and neuroscience to investigate data summarization techniques for diverse collections of images.

Requirements:

Candidates should be pursuing an undergraduate or graduate degree in computer science, mathematics, statistics, electrical engineering, or a related field.

Desired Skills:

Candidates should have a background in machine learning, data mining and/or statistics and be familiar with computer vision literature on feature extraction. The candidates should have a strong grasp of a high level programming language (e.g. Python or C++). Finally, candidates should be interested in designing experiments and exploring state-of-the-art techniques in computer vision.

Keywords:

Computer vision, machine learning, data mining, big data and deep learning

*Summer Research Program Intern in Cognitive Informatics, Requisition #2563***Project Description:**

A summer intern is needed to provide support to an ongoing project exploring human performance in the intelligence gathering field. Of particular interest are the cognitive processes involved in the intelligence gathering, synthesizing and reporting processes. The successful applicant will participate in human-subjects research, supporting the refinement of the experimental procedure, upgrades to the

experimental apparatus, data collection from participants, data recording and analysis. Of particular interest will be recording and processing of eye tracking data for assessing the comprehension of intelligence reports and cognitive load based on pupil measurements.

Requirements:

Applicants must have an educational background in cognitive science or a related discipline (e.g., experimental psychology, human-computer interaction, or human factors engineering), with experience conducting human-subject research. Experience with eye tracking equipment and other physiological recording methods are desired. The successful applicant will be able to support the principal investigators during testing in all phases of research, including knowledge elicitation, experimental design and setup, subject preparation, data collection, data management and analysis, and report generation

Send resume to:

kelly.geyer@ll.mit.edu