Research and Project Experience

Chirag Nagpal

During the summer of 2014, I received the prestigious Science Academies' Research Fellowship to pursue undergraduate research under the guidance of Prof. N. Balakrishnan, Associate Director, Indian Institute of Science (http://www.serc.iisc.ernet.in/~balki/) at the Supercomputer Education Research Centre. During this time I worked as part of the 'Social Network Analysis Group'. Here I worked on a project entitled 'Application of Machine Learning Techniques for Twitter User Classification'. The project involved the application of Decision Trees, Bayesian Classifiers and Vector Machines in order to develop classifiers to classify Twitter users based on their interests, into Sports, Music and Politics. Python scripts were developed to aggregate Twitter User Profile information using their REST API. Text preprocessing, including stemming, and stop word removal was carried out. To implement the classifiers, Python packages NLTK and SciKit were utilised. The implemented classifiers had accuracies of over 80% in four cross validation. Complete report from the project is available at http://chiragnagpal.com/finalreport.pdf and has also been consolidated into a paper 'Twitter User Classification using Ambient Metadata', available on the arXiv (http://arxiv.org/abs/1407.8499)

As part of my college Big Data Research Group (http://ait-big-data-initiative.github.io/), I have implemented a project titled 'Spatio-Topical Modelling of Tweets from India. The project involves, the analysis of information dissemination through Twitter in the Indian subcontinent. Over 400,000 geo-referenced tweets were aggregated from Twitter. Document-Term-Matrices were generated for the corpus, with TF-IDF weighting and unsupervised learning algorithm, Latent Dirichlet Allocation was applied to the corpus, in order to discover trends and topically model the tweets. They were then mapped according to their latitude longitude cooridinates from the georeferenced data. This project was implemented in R. R Packages like 'tm', 'topicmodels' & 'mapdata' were utilised.

Currently, I am working on a deep learning recommending algorithm for friends on facebook. The algorithm, recursively iterates through the pages liked by a user, creating a 'like tree', intersections between the 'like trees' of various users are then utilised in order to reccommend users for friendships.

Apart from my project work in Data Analytics and Machine Learning I have interest in Ubiquitous Computing and the Internet of Things, and some of my projects from this domain have been featured in leading blogs like **'Hack A Day'**[1][2] and '**Dangerous Prototypes**'[3][4].

A complete list of my projects is available at: http://chiragnagpal.com/projects.html

My CV is available at: http://chiragnagpal.com/cv.pdf

1 http://hackaday.com/2014/04/06/hackaday-links-april-6-2014/
2 http://hackaday.com/2013/12/07/speeding-up-beaglebone-black-gpio-a-thousand-times/
3 http://dangerousprototypes.com/2014/03/31/beaglebone-black-and-msp430-based-wireless-sensing/
4 http://dangerousprototypes.com/2013/12/18/laser-pointer-controlled-with-a-pc-mouse/01111