Raj Balasubramanian schroskat@ieee.org

RESEARCH STATEMENT

My interests are in the intersection of software engineering and the domain of health care. I am interested in applying software engineering practices and methods and its application to the domain of health care. Specifically I am interested in applying the wide range of tools and techniques available from the domain of software engineering to the existing mathematical methods and modeling tools available to the economist, ecologist and epidemiologist. I believe that by accomplishing this, it will provide the decision makers and scientists not only with better understanding of the underlying phenomenon and alternative forms to view the underlying processes but also provide “good-enough” estimates in realtime to make meaningful decisions. The topics in health care that I am interested in range from understanding the basic disease dynamics of communicable diseases to the allocation of resources in a given health care provider or a region to better handle disease outbreaks. Specifically I am interested in augmenting and enriching existing disease models with underlying topological, demographical and geographical structures that are prevalent amongst the communities where the disease spreads. I also intend to leverage the public web and other social structures available from the web to augment the original health data. Therefore, by capturing some of these basic properties of the community, the overall disease model (both in inference, prediction and control) become more meaningful to the local and regional authorities.
In order to pursue this research I am structuring few key projects that I will work in order to complete my dissertation in the near-term and then outline further areas of exploration for future work. In the short-term, I am working on three projects-

- Hospital Discharges in Texas are on the order of million patients a quarter and these records are made public by the state. The first derivative from this dataset is to characterize human mobility patterns for healthcare and to understand the county to county dynamics for patient travel as opposed to mobility patterns for work. The second derivative from this dataset is to better understand and characterize dynamics of influenza-like illnesses (ILI) at the regional and local levels while corroborating with publicly available data from Google Flu Trends and the CDC ILINet. I will be applying network modeling and the software from the graph mining literature to characterize and study the mobility patterns. I will also leverage Map-Reduce paradigm of mining the large datasets in the UT computer clusters. I will be looking at better ways to store, replicate, visualize and present the movement data both spatially and temporally.
- Cholera is an old disease that has been known since the time of Greeks and ancient India. It has been studied extensively and modeled in the past century. But inspite of this, there has been a growing chasm between the how the localist and the contagionists have explained the outbreak of cholera. Although, there has been some recent work in taking into account the climate and geography of the region in explaining the outbreaks, there is no clear global model that is simple enough to qualify the outbreaks taking into account the local variance. My intent here is to have a global model for cholera using bipartite graphs as the underlying structure and apply network based dynamic models to characterize the disease spread and outbreaks. I intend to paramteirze my models from the historical data from Bengal (from colonial India) as well as from the recent outbreaks in Bangladesh, Haiti, Zimbabwe and the military occupied Iraq and Afganisthan.
- Characterizing networks – weighted directed graph structures. This project

has two aspects. In the first subproject - I will look at using formal methods from software engineering to create large graphs with desired properties as well to model disease phenomenan using petri-nets in aiding better simulation of the models. In the second sub-project I will catalog various properties of directed weighted networks from social web, transportation network, human work mobility network, email and blog networks and investigate key features that characterize the network and the process yielding the network. This will involve creating a body of software framework from the available graph mining and network modeling packages as well as finding ways to scale that software framework across cluster of servers.

Educational & Employment Summary

University of Texas at Austin

◦ PhD candidate – Software Engineering (expected graduation Fall 2012). Co-supervised by Dr. Lauren Meyers (Integrative Biology) and Dr. Joydeep Ghosh (Electrical & Computer Engineering)

◦ Master of Science in Engineering - Software Engineering, 2008 (GPA: 4.0)

IBM

◦ Enterprise Solution Architect, Software Group, 1997-current

---

Tools & Techniques

I bring over 15 years of IT work experience from a variety of industry domains and background in solving the business problems in these domains using solutions developed using software engineering principles. For the near term research projects I plan on using the following tools and techniques -

Languages – My primary language for developing new software will be based on Python, C and R (Statitical Software). I will also leverage Java where packages exist that already solve the problems – for example, I will leverage STEM in initial epidemiological modeling with GIS support. I will also investigate the use of Haskell in creating a Domain Specific Language (DSL).

Storage – I am exploring the use of non-relational databases, such as Hbase, CouchDB and MongoDB for storing, replicating and processing large datasets.

Data Processing – I will use Hadoop based Map-Reduce infrasturcture and related projects from Apache Hadoop project to mine and learn from these large datasets.

Application Framework – I will look at creating an application framework using REST consturcts in exposing any realtime access (by human users) to various analysis and learned information via standard web mashups using Google Maps and visualization engines using Flex.

References

1. Google Flutrends - http://www.google.org/flutrends/
2. Apache Hadoop - http://hadoop.apache.org
3. Meyers Lab - http://www.bio.utexas.edu/research/meyers/
4. IDEAL - http://www.ideal.ece.utexas.edu/
5. Eclipse STEM project - http://www.eclipse.org/stem/

Last Updated: January 2011