



Dongbo Hu <dongbo.hu@gmail.com>

Penn data scientist position

Amol Navathe <amol@mail.med.upenn.edu>

Mon, Aug 24, 2015 at 1:13 PM

To: Dongbo Hu <dongbo.hu@gmail.com>

Hi Dongbo,

Thanks for your prompt reply. We are starting a large new initiative in collaboration with the University of Pennsylvania Health system. The goal is to move medicine from a more reactive paradigm to one that anticipates and treats clinical events before they occur. I am including a summary below plus example qualifications. If this seems of interest, please let me know right away and we can set up a time to talk on the phone or meet in person if you are close by.

Best wishes,

Amol

SUMMARY OF THE INITIATIVE

Objective: To improve the health of Pennsylvanians at an individual, community, and population level by changing the paradigm of medical care and health care delivery to predicting and preventing onset, exacerbation, and advancement of disease rather than the current standard paradigm of treating patients after they experience clinical events. The prospect of 'big data' making improved health and health care possible is enticing, but to maximize impact in these contexts requires not just the application of cutting-edge techniques but *also* an essential understanding of human intuition and clinical knowledge. For instance, we tend to overweight the likelihood of both oft-discussed and recently witnessed outcomes; overweight small probabilities; exhibit overconfidence in the precision of our estimates; detect patterns in data where there is, in fact, only noise; and neglect base rates. Because medicine is so reliant on human judgment, the behavioral economics literature suggests that there is a significant opportunity to (a) reduce variance in and (b) improve the quality of the medical care delivered to diverse populations if systematic errors in judgment could be prevented by using a combination of human and machine intelligence to make better prediction algorithms and decisions. We will work to develop better early warning systems, better decision support, and better algorithms to predict readmissions and chronic disease exacerbations before they happen to facilitate better targeting of interventions to improve the health of Pennsylvanians and reduce disparities. These will include

health system interventions and population level interactive tools to engage individual citizens (not just patients) in predicting clinical, epidemiological, and behavioral outcomes.

Specific Aim #1: Develop algorithms to predict in-hospital clinical events and disease progression in specific patient populations

Sample Project: Develop an application using both structured and unstructured data to dynamically monitor and predict vascular surgery patients at risk for post-surgical complications to be used at the bedside. Algorithms would leverage multiple types of data across health systems (e.g. EMR and real-time monitoring data), payers (e.g. claims), and state resources (e.g. Medicaid) within Pennsylvania.

Specific Aim #2: Utilize data from individuals who are not hospitalized (such as wearable devices) to supplement clinical and claims data to predict individual clinical events such as hospital readmissions and chronic disease exacerbations.

Sample Project: Develop an application to incorporate wearable device data into risk models for congestive heart failure exacerbations and hospital readmissions.

Sample Project: Create algorithms to monitor patients for depression after Coronary Artery Bypass Graft surgery based on individualized data drawn from devices and Pennsylvania health providers across the continuum of care (e.g. cardiac rehabilitation).

Specific Aim #3: Connect online data from social media and other sources to track the health of Pennsylvanians in an interactive and transparent way.

Sample Project: Develop a publically accessible web application which extracts and generates dashboards from data on social media sites (e.g. Twitter, Yelp, Facebook) about cardiovascular health (e.g. symptoms, diet, exercise, medications, healthcare facility resources) merged with structured data (e.g. PHC4 hospital data, HCAPS, census data, Aetna claims [from HCCI], Medicaid claims) for region specific surveillance, real-time resource allocation, and measurement of cardiovascular health outcomes.

QUALIFICATIONS

Penn Medicine is seeking a Data Scientist to join our Data Science team. At its core, this position is about revolutionizes healthcare with data driven analytic methods. The position will apply machine learning and computational statistics techniques to clinical, administrative, and social media data to create immense value for patients. Individuals in this role are expected to be comfortable working as a software engineer and a quantitative researcher. This position will develop state-of-the-art algorithms, tools, and apps. Specifically, programming will range

across a subset of text-mining, machine learning, natural language processing, distributed computing, website development, mobile app development, computational psychology, and data visualizations.

Responsibilities:

- Work closely with clinicians and administrators to identify and answer important healthcare questions with appropriate statistical techniques on available data.
- Lead large data acquisition, data mining, and analysis techniques to time series, NLP, imaging, structured clinical data and social media websites.
- Create and manage protocols for the analysis of passively collected information
- Responsible for the design and execution of research and development projects utilizing advanced computational techniques to generate novel medical insights and to build clinical decision support systems.
- Communicate findings to clinicians and administrators
- Drive the collection of new data and the refinement of existing data sources
- Work closely with a team of developers and researchers to conduct scholarly work at the intersection of social media and health
- Design and create the necessary algorithms to facilitate research, including either: (a) text and data analysis code or (b) web and mobile app code for acquiring data and disseminating results.
- Implement algorithms (write code) to facilitate research or disseminate results in either Python, Java, Matlab, Javascript, or C++.
- Examine and maintain data in MySQL and NoSQL style databases.
- Perform technical literature reviews, including identifying relevant papers in computer science and writing accessible summaries for the team.

QUALIFICATIONS:

- M.S. required in a relevant technical field (e.g. statistics, mathematics, physics, computer science)
- At least 3 years of experience required in a relevant role and at least 2 years of experience solving analytical problems using quantitative approaches required
- Fluency with statistical and machine learning techniques such as decision trees, neural networks, graphical models, clustering, dimension reduction, and deep generative models required.
- Experience manipulating and analyzing complex, large volume, high-dimensionality data from varying sources required

- Fluency in coding in C++, Java, or Python required
- Fluency in MySQL, NoSql design, deployment and analyze on many TB's of data preferred
- Experience not required but preferred in: Hadoop/MapReduce, text-mining, natural language processing, machine learning, apps
- A strong passion for empirical research and for answering hard questions with data.
- A flexible analytic approach that allows for results at varying levels of precision
- Ability to communicate complex quantitative analysis in a clear, precise, and actionable manner
- Familiarity with relational databases and SQL preferred, Linux
- Candidates will have demonstrated the ability to work as part of a team as well as independently; have experience working with patients and guarding sensitive information.
- Applicants must be able to multi-task, attend to detail and maintain a high level of accuracy.
- Prior experience supervising/training others preferred

From: Dongbo Hu [mailto:dongbo.hu@gmail.com]

Sent: Monday, August 24, 2015 7:28 AM

To: Amol Navathe

Subject: Re: Penn data scientist position

[Quoted text hidden]