Dear Hiring Manager,

I am writing this letter to express my interest in the research engineer position posted on the career website of your institution. I recently obtained my PhD degree in Computer Science at Hong Kong University of Science and technology. My Advisor is Prof. Huamin Qu. During the year 2014, I was also a visiting student at Georgia Institute of Technology.

During my PhD study, I developed visual analytics applications that combine data mining and interactive visualization techniques to facilitate exploration and knowledge discovery in multivariate graph data and collections of social media data. I will briefly introduce some of my research projects and industrial projects bellow. Besides that, I will also introduce my technical background and describe my past experience and my interest in open source software development.

▬ Research Background

Visual analysis of social media data

Social media data is representative of the quantity and the complexity of the data that we are facing today, and many research projects have been carried out to unravel the various social process underlying the stream of posts. In this project, I developed a visual analytics system to help the domain experts study how various topics compete to attract public attention when they spread on social media, and what roles do opinion leaders such as mass media, political figures and grassroots play in the rise and fall of various topics. The system combines text mining, time series modeling and interactive visualization to help gain insight into the temporal dynamics of topic competition on social media (e.g.,Twitter).

Multivariate graph data analysis

I proposed and implemented an interactive visualization system which incorporates co-clustering algorithms to facilitate the identification and exploration of clusters formed by nodes with common connections in a bipartite graph. The system also supports correlation analysis between the node groupings and the features of the nodes. The visualization framework is generally applicable to any data that can be modeled as a bipartite graph. One data set I used in the case study is the roll call vote records of the US Senators on the passage of bills. The system facilitates the analysis of voting pattern of the senators and its correlation with their party affiliations and the states they represent. It also helps detect groups of bills with similar sets of supporters.

Future directions

I believe that to harness the value of big data in the future, we need to create a closer bond among the areas of data mining, visualization, and human computer interaction. I will continue to explore the design space with different possible combinations and develop systems that could effectively help the analyst to have the best from each.

▬ Industrial Collaboration

I have worked closely with partners from industry in several projects, building visual analytic systems that can be combined with their data storage infrastruture and help analyze various data sets such as mobile broadband communication records and mobile user checkin data. The systems are built to inform business development with human mobility and communication patterns. We make design decisions by identifying their requirements, and potential areas for innovation. In some projects, I lead the design of the system architecture and coordinate the development efforts in our group.

▬ Technical Background

In the past, I have used a lot of open source libraries to prepare and analyze data, to develop interactive visualizations, and to implement research prototypes. These include D3.js, Three.js, networkx, numpy, scikit-learn, igraph, Prefuse and so forth. Based on my own experience and observation, I believe the development of these libraries has greatly benefited and even transformed the research activities across a wide range of disciplines, and I hope to be a part of it.

Recently, I have published some open source project including a library in javascript for the creation, manipulation, and analysis of graphs as part of a research prototype, and a library that supports fast in browser density map generation for one million points in real time by harnessing the parallel processing power of GPU. They can be found on my github page (http://github.com/lliquid). I also plan to become more engaged in the open source community in the future.

From another point of view, as an information visualization researcher, I often consider one role of myself as a toolsmith who designs software applications and libraries that can lower the barrier and reduce the effort for the data enthusiasts and domain experts to create informative visualizations, and discover knowledge from the data they collected. I am also willing to contribute to the overall process of crafting the code, documenting apis, drafting tutorials, and promoting good practices. It is with great interest that I submit this letter as an application for the research engineer position.

Sincerely yours,

Panpan Xu