Digital Science: Catalyst Grant Proposal

Title of the Proposed Project: Human-machine analytics for visual recommendation

Point of Contact:

Tommy Dang, Assistant Professor
Computer Science Department
Texas Tech University Lubbock, Texas 79409-3104
Phone: (+1) 806-319-3156
Email: tommy.dang@ttu.edu

Prepared date:

December 25th, 2018

---

1. ABSTRACT

Limited human cognitive load and finite computing resources, display resolutions are the major obstacles for developing interactive visualization systems in large-scale data analysis. Our strategy to bridge these gaps is to present the *right* amount of information to the *correct* users through the use of *effective* visualizations. Therefore, the first step in this process is to capture users' interests, background, reasoning process, and cognitive style. A straightforward approach is to silently record user interactions with the visual interface since this does not require extra effort from the user side. Then the next questions are how to infer user aspects from their interactions and how to translate them into effective visual representations. This work proposes a systematic approach to formalize user interactions and visualizations through the use of declarative grammars of visualization specification to enhance the symbiotic relationship between human and computer through the development of interactive systems that can sense and learn the human perceptual and cognitive states while possessing the ability to adapt their visual interest.

2. TEAM

PI Tommy Dang is an Assistant Professor of Computer Science at Texas Tech University where he directs the interactive Data Visualization Lab (iDVL). His research on big data visualization and visual analytics has appeared in Computer Graphics Forum and IEEE Transactions on Visualization and Computer Graphics and presented at IEEE Information Visualization among others. His ongoing research focuses on novel combinations of visual representations. The PI's ambition is to effectively connect the three components: Big Data-Visual Interface-Users. http://myweb.ttu.edu/tnhondan/iDVL/index.php

Developer Ngan Vuong: Ngan earned her Bachelor's Degree at University of Science, Vietnam in 2014. Her interests are in data science, user interface, and signal processing. Ngan joined iDVL in Fall 2018 and working toward her PhD under the supervision of **PI Tommy Dang**.

3. PROBLEM & SOLUTION

Visualizations are context-specific. To understand the context of visualizations before deciding to use them is a daunting task since users have various expertise/background and there are thousands of available visual representations (and their variances). To this end, the team will develop a visual analytics framework to archive the following goals (i) to automatically generate a number of suitable representations for visualizing the input data and present to users as a catalog of visualizations with different levels of complexity (ii) to infer aspects of the user's interest, background knowledge, reasoning

process, and cognitive style based on the user's interactions using the state of the art data mining techniques (iii) to narrow down a smaller set of visualizations that suit users analysis intention. The results of this process will give our analytics system the means to better understand the users' analysis process and will enable it to better provide timely recommendations.

The aim of the project is to reduce the cognitive workload of the analyst and move the user interaction into the visual space. If successful, the proposed work allows expert users to quickly narrow down to the right visual representations for their analysis and at the same time allows novices to open a broader range of visualization to novices and helps them find the right visualizations without the need to develop separate expertise.

4. PRODUCT

We develop a visual analytics framework automatic generation of visual representation which can learn and improve the recommendation performance from user interactions. Our automated visualization framework is designed based on the following considerations: (1) how to handle large, high-dimensional, and complex (various data types) data (2) how to balance automation and user control. Here is the link to our prototype: https://idatavisualizationlab.github.io/N/HMAviz/index.html

Figure 1: Our visual interface: (1) Overview, (2) Exemplar, (3) Navigation, (4) Focus, and (5) Expanded view.

Our approach silently captures user's interest on the following dimensions: visual abstraction levels, visual features (such as outliers, clusters, and correlation), and the number of variables to be plotted. The following figure provides our conceptual catalog of visualizations projected on these dimensions on the chemical concentration data collected from the PXRF.

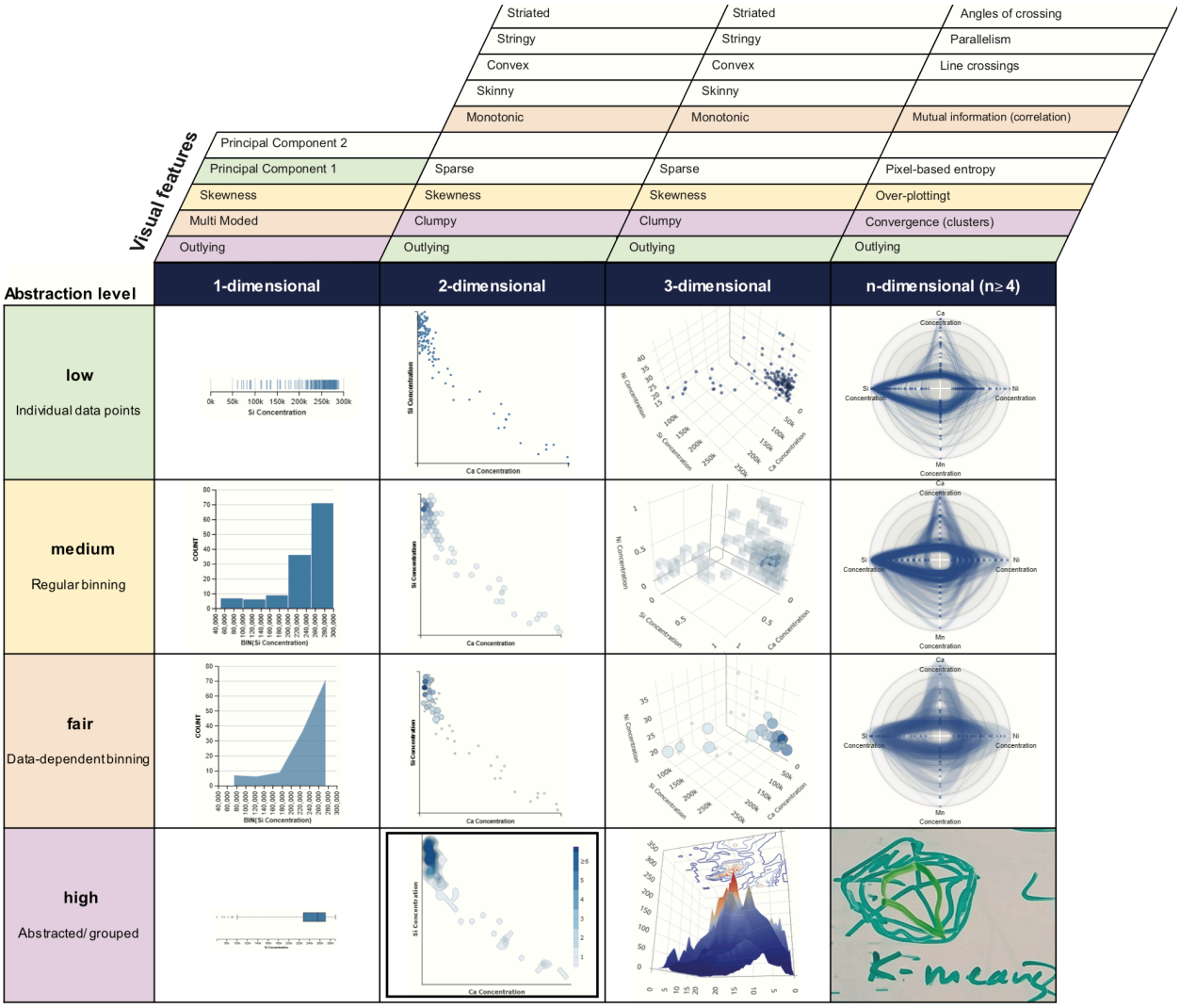

Figure 2: *Our conceptual visualization catalog projected on abstraction levels, visual features, and number of dimensions.*

We are evaluating the product for interactive and exploratory data analysis of two classes of populations: novices and experts. We first selected to work on the chemical measurement data produced by the PXRF since companies are moving the data to the clouds and there are substantial needs to provide real-time visual analysis of these data to customers (based on our 29 customer interviews). Besides that, we have academic collaborators who are interested in using our software to speed up their everyday tasks.

5. PRICING

Our approach captures historical data for each user account (similar to Netflix records which movies that you watched along with your predefined preferences) and therefore provide timely recommendations. For example, the visual interface can warn users about possible errors or outliers, instantly on-field for further investigation when the users are still on-field. Mistakes in data collection via the PXRF can be corrected with the lowest cost. We are going to charge a monthly subscription fee for the individual user account or yearly /lifetime license to companies for their customer usage.

6. LONG TERM VISION

Multiple data types are more challenging and require a systematic approach. To handle many complex real-world scenarios (the input data may contain images, text, relationships, and geographic locations), we **plan to** use declarative languages for multi data type visualization. *Figure 3* shows the hybrid of two popular visualizations where time series, numeric, text, and categorical data are visualized in a single picture: https://github.com/iDataVisualizationLab/CVSS

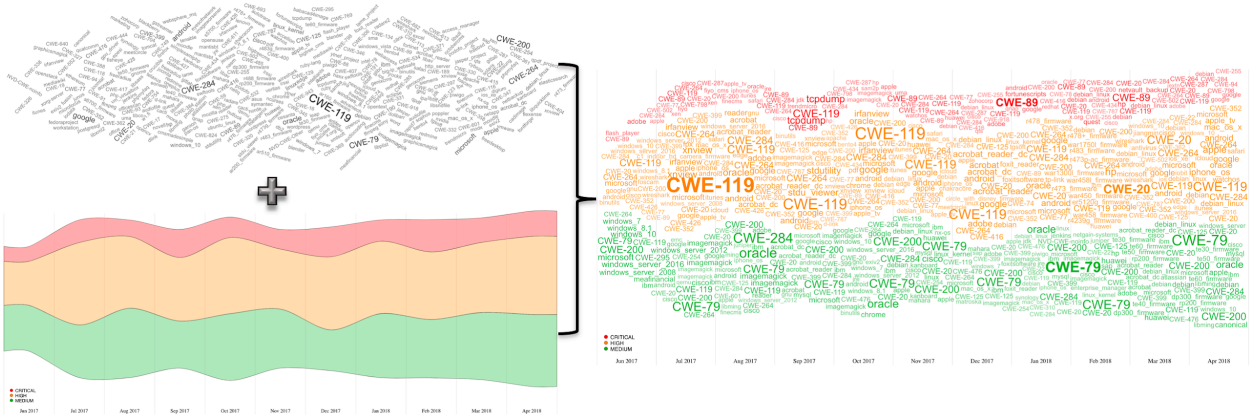

Figure 3: Popular word cloud and streamgraph are systhesized into snapshot of different data types: Time, numeric (severity levels are displayed vertically: medium, high, and critical), and text .

7. COMPETITORS

The knowledge discovery process is time-consuming and might be handled by a team with different expertise. For example in soil analysis, the entire process can take 3 days to a week. After on-field data collection, soil scientists analyze the chemical measurement data on lab computers using custom software, or excel, or eyeball the data. Once an issue occurred, it is difficult to get back to the field to re-sample the data. This is a cumbersome process and sometimes impossible. Our approach aims to simplify this into a one-step process: recommending the real-time visual representations accounting for user education and preference right on the user cell phone/tablets. As we first focus on the visual recommendation for chemical measurement, we found the following competitors:

Reference	Name	Description	Date	Relevant Problem	Inventor
Patent 1	X-Ray Fluorescence Analyzer	System and method for processing hand-held XRF data.	Publication: Jan 31, 2018	Reduces constructive x-ray reflection peaks and provides on-site analysis of raw data.	Bruker AXS Handheld, Inc.
Patent 2	X-Ray Fluorescence Visualizing, Imaging, or Information Providing of Chemicals, or Compounds,	Method for determining chemical species using PXRF and visualizing the analyzed data for the user.	Publication: Apr 2, 2009	Improves the available visualizing of the information gathered from the data.	Searete LLC
Patent 3	Method and System for Obtaining Knowledge Based Recommendations	A method and system for obtaining knowledge recommendations. A request with multiple input parameters is received for a desired outcome.	Publication: Mar 14, 2002	Improves identification, selection, validation and screening of compounds.	Cellomics, Inc.

Patent 1 allows users to interact with the device (using the filter wheel) to define a new spectrum range then update the display to this new spectrum data. In other words, they provide data filtering and refinement on specific visualizations (as curves on a graph), not recommend a more suitable set of visualizations accounting for user education and preference. Moreover, our approach captures historical data for each user account (similar to *Netflix* records which movies that you watched along with your predefined preferences) and therefore provide timely recommendations. Mistakes in data collection via the PXRF can be corrected with the lowest cost.

Patent 2 only seeks to apply this to medical applications and does not mention anything about recommending visualizations according to the user's knowledge in the field. Our applications are much broader and accounting for user education and preference.

Patent 3 simply provides a quick "*yes* or *no*" type result. In our approach, novices can start their data exploration process with the first column in ***Figure 2*** while soil scientists are more interested in chemical correlations and spatial contour map (the last cell in the second column of ***Figure 2***). By simply adding color encoding to the contour map, our framework can generate the following concentration heatmap for a selected chemical element: https://idatavisualizationlab.github.io/Soil/

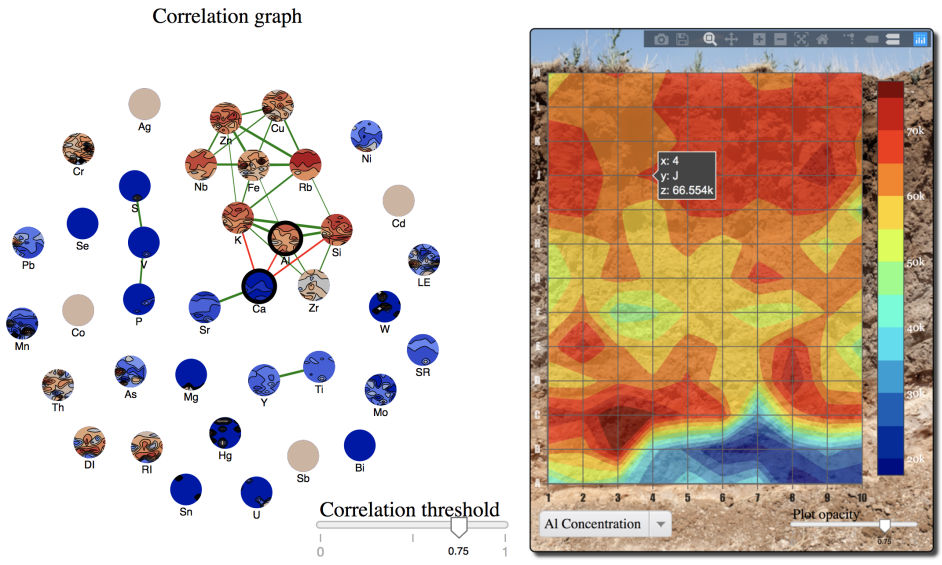

Figure 2: *The generated heatmap for Al concentration.*

8. MARKET

Time to Market	Greater than 36 months	18 to 36 months	Less than 18 months
Investment to Market	Greater than $750,000	$250,000 to 750,000	Less than $250,000
Potential Market	Limited	Medium	Large,m

For the chemical measurement analysis for the PXRF, we have the following potential commercial partners. We can acquire new customers from other applications as well.

Name	Contact Information
Olympus	Dr. Michael Hull, 110 Magellan Circle | Webster, TX 77598. Phone: (713) 485-9915. Email: michael.hull@olympus-ossa.com
Hitachi	Mr. Rose Jordan, 2 Technology Park Drive, Westford, MA 01886. Phone: (978) 850-5560. Email: Jordan.Rose@hitachi-hightech-as.com

9. PROGRESS TO DATE

We already interviewed 29 customers and also developed one prototype software. Our next immediate step steps are to extend to other measurement devices and domains.

We have filed a patent. Here is the technology assessment report on December 12, 2018: "Based on a preliminary analysis of the relevant literature, the disclosed technology **is patentable**. The inventor proposes that the visualization aspects are superior to technologies."