Eurographics Conference on Visualization (EuroVis), Posters Track (2016)
T. Isenberg and F. Sadlo (Guest Editors)

MetricScalpel: Analyzing Diagnostic Outcomes with Exploratory Data Visualization

W. Li[1], J. Elderman[1], E. Hoogerwoord[1], M. Funk[1], A. C. Brombacher[1]

[1]Eindhoven University of Technology, Netherlands

Abstract

Healthcare data is the emerging force to push ahead smarter clinical solutions. However, unforeseen challenges like unmanaged massive data and costly access to it make it difficult for domain experts to easily derive actionable insights. In collaboration with a local diagnostic service provider, we designed MetricScalpel, a web-based visualization tool to help people quickly look into the real-life diagnostic data in the local community. The tool enables swift exploration into the multivariate diagnostic data sets from different sources and facilitates data selection/subsetting for deeper analysis. It can be used to reveal overlooked health conditions on almost any level without the requirement of heavy technical knowledge. Such design makes it easier to be accepted by a wider user group in the healthcare related organizations. It was proven to well serve the domain experts in validating pre-exist hypotheses in cohort analysis as well as revealing undiscovered patterns of health conditions in the local community. External evaluation shows operation cost was remarkably confined as domain experts were assisted with direct and intuitive access to the relevant data in need.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

Applications based on *electronic health record* (EHR) information have drastically proved its value in clinical decision making as informatics is expanding its significance in healthcare [SSMT08, LPZ15, BvGGS16, MSC*04, FOH*06]. Whether it's manual retrieval or aided by advanced techniques like data mining, the benefit generated from healthcare data should not be overlooked [AC12, NWSL12, ZAS*13, SSB*12, ABR*15, Saa10]. However, to realize the full value of healthcare data in a cost effective manner, more integrated procedures should be implemented in the cluttered data analysis process [KPHH12]. It should also empower the domain experts to explore and communicate complex patiences' diagnostic data sets (as an aggregation from different data sources) with ease [MDM*15, ZGP15].

In the context of this project three major roles are identified: the patients who contribute the samples to be tested, a number of clinical institutes scattered around the local community who collect the sample, a centralized diagnostic service provider who conduct the diagnostic analysis and report the outcomes. Being the main partner we work with, the diagnostic service provider has the capacity to conduct assays to spot a broad range of diseases and store the outcome in a digitalized format. As their customers, external institutes send blood or urine samples to diagnose and wait for the test result. As the diagnostic data are centrally managed and shared (under restrictions) among third parties, the data became a valuable asset for many professionals both internally and externally though substantially under-exploited currently. Little support has been provided to allow the knowledge obtained by analysis in the laboratory to be managed by a central ICT system and easily communicated with all the three roles (or at least between people in the diagnostic service and clinical institutes).

2. Design of MetricScalpel

Considering urine assay data sets are temporal organized and include a long time period, a longer timeline was chosen to allow precise selection to a day level. Since comparison capability is of more urgent need for domain experts, we grounded the design with a juxtaposition layout to support this. As a result, quick comparison of precisely defined metrics of urine assay data sets between user defined patient group is made possible, Figure 1. Further filtering upon external institutes can be realized by clicking the institutes' ID on the left-most menu list if required by certain scenarios.

A smaller interactive histogram in the upper center subsection indicate detailed distribution of the chemical metrics, combined with a bar chart placed on top showing the pH value. Donut charts on the upper right corner indicates the percentage of positive/negative test result of the few critical chemical components. To better suggest the three states of healthy (green), neutral (grey) and unhealthy (red), consistency of color choice for aesthetics here is sacrificed for better usability .

DOI: 10.2312/eurp.20161129

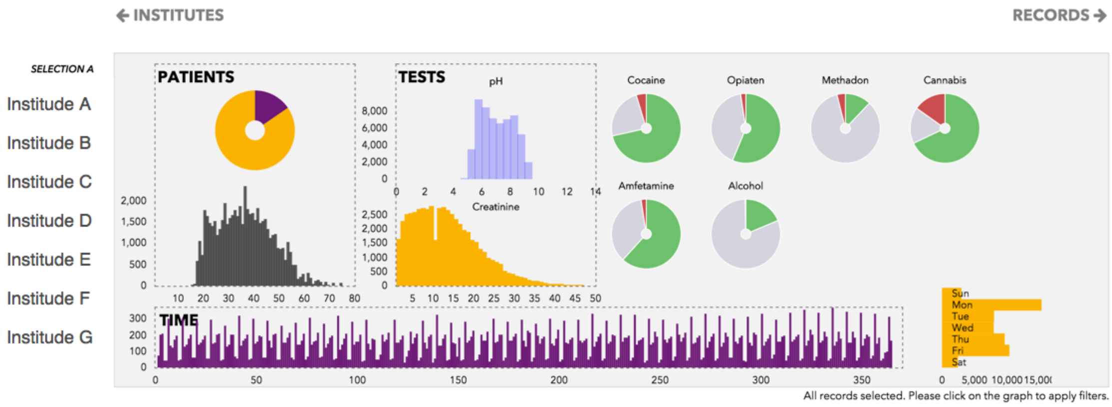

Figure 1: *The design consists of two near-identical panels like above, which are vertically aligned to compose a juxtaposition view. The "INSTITUTES" menu list is used to fold/unfold the institute list. The "SELECTION A" area is in the default state when no filtering is applied. Selection can be made upon itself or the following "SELECTION B". This part is designed to show how the institute differs from the rest by referring to two different selections. "RECORDS" can be clicked to show the raw data responding to selection.*

The "RECORDS" panel on the right side of the screen is designed for raw data export. User can target on a subset group of data which may contain anomalies or promising trends by selecting the desired data range and use this for further research.

3. Interaction & Usage

The tool is open-ended and designed with adaptability in mind. Use case can vary from user to user. The followings are a few typical usage structured under the input from the potential users in the diagnostic service provider.

Cohort Analysis:Cohort analysis involves determining the independent contribution of a variety of factors [SZCDMG*16]. Group filtering functionality is used to find out the effect of an isolated factor. Here, filtering is supported by flexible selections. And quick fine adjustment on several features is facilitated to discover the cohort characteristic in detail. Specifically, user is allowed to cross-filter upon demographic distribution, temporal range, age range, day of the week. So theoretically, determining factors in these variables are made possible. After filtering and zooming into the target diagnostic data are finished. It is possible to export the selected data into a well structured format for more in-depth cohort analysis. This is useful when stakeholders with different interests and backgrounds may seek to scrutinize the data in their own preferred ways.

Institute Benchmark: One common tedious task before the tool was devised is experts in the service provider's responsibility to answer to the institute's inquiry regarding their test outcome and observe how it varies on a macro-perspective level. With the tool at hand, now we can do this easily by defining the selection to certain institute and perform research. For example, by targeting at the same institute on the left panel, two selections of age range of the same length can be made on both sections respectively. Consequently, how testee number changes in season is presented. Moreover, institutes are located in the same region and can be geo-coded, meaning comparing the positive report rate between them may also be facilitated to reveal certain clinical patterns geographically.

4. Evaluation & Future Work

User feedback shows that the tool not only improved access to data with exploratory interactivity and multi-dimension manipulation in parallel, it also helped internal experts as well as external parties quickly dive into the data sets and conduct insightful research more independently than before. Significant reduction of operational cost is validated as a result of the streamlined procedures involved. For instance, they are now able to spot an anomaly by making comparisons in one scale and trace it with a few clicks immediately without switching and complex reconfiguration. The tool also significantly reduced the time or effort in asking for assistance back and forth in the organization.

Further improvements can be as follows: 1) facilitate better integration with the database and reduce the latency between data used in visualization and data collected in real time. 2) Improve the usability by tweaking UI elements and optimize the performance. 3) include the geo-location of external parties, which is available in the dataset yet was not put into practical use, e.g., for further researching a potential impact of transportation distance on urine sample quality.

References

[ABR*15] Ancker J. S., Brenner S., Richardson J. E., Silver M., Kaushal R.: Trends in public perceptions of electronic health records during early years of meaningful use. *The American journal of managed care 21*, 8 (2015), e487–93. 1

[AC12] AbuKhousa E., Campbell P.: Predictive data mining to support clinical decisions: An overview of heart disease prediction systems. In *2012 International Conference on Innovations in Information Technology (IIT)* (2012), IEEE, pp. 267–272. 1

[BvGGS16] Beentjes T. A. A., van Gaal B. G. I., Goossens P.

J. J., SCHOONHOVEN L.: Development of an e-supported illness management and recovery programme for consumers with severe mental illness using intervention mapping, and design of an early cluster randomized controlled trial. *BMC health services research 16*, 1 (2016), 20. 1

[FOH*06] FUNG V., ORTIZ E., HUANG J., FIREMAN B., MILLER R., SELBY J. V., HSU J.: Early Experiences with e-Health Services (1999-2002): Promise, Reality, and Implications. *Medical Care 44*, 5 (May 2006), 491–496. 1

[KPHH12] KANDEL S., PAEPCKE A., HELLERSTEIN J. M., HEER J.: Enterprise Data Analysis and Visualization: An Interview Study. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2917–2926. 1

[LPZ15] LIU W., PARK E. K., ZHU S. S.: e-Health security context management service. In *2015 International Conference on Computing, Networking and Communications (ICNC)* (2015), IEEE, pp. 379–383. 1

[MDM*15] MALIK S., DU F., MONROE M., ONUKWUGHA E., PLAISANT C., SHNEIDERMAN B.: Cohort Comparison of Event Sequences with Balanced Integration of Visual Analytics and Statistics. In *the 20th International Conference* (New York, New York, USA, 2015), ACM Press, pp. 38–49. 1

[MSC*04] MURPHY J., STRAMER K., CLAMP S., GRUBB P., GOSLAND J., DAVIS S.: Health informatics education for clinicians and managers–what's holding up progress? *International Journal of Medical Informatics 73*, 2 (Mar. 2004), 205–213. 1

[NWSL12] NING Z., WENXING H., SITING Z., LI T.: A solution for an application of information visualization in telemedicine. In *2012 7th International Conference on Computer Science & Education (ICCSE 2012)* (2012), IEEE, pp. 407–411. 1

[Saa10] SAADE J.-H.: Use of the Electronic Health Record (EHR) to Optimize Communication of Isolation Practices. *American Journal of Infection Control 38*, 5 (June 2010), e129–e130. 1

[SSB*12] SINHA P., SUNDER G., BENDALE P., MANTRI M., DANDE A.: USA's EHR Meaningful Use. In *Electronic Health Record*. John Wiley & Sons, Inc., Hoboken, NJ, USA, Dec. 2012, pp. 295–309. 1

[SSMT08] SULAIMAN R., SHARMA D., MA W., TRAN D.: A Security Architecture for e-Health Services. In *2008 10th International Conference on Advanced Communication Technology* (2008), IEEE, pp. 999–1004. 1

[SZCDMG*16] SIERRA ZÚÑIGA M. F., CASTRO DELGADO O. E., MERCHÁN-GALVIS A. M., CAICEDO J. C. C., CALVACHE J. A., DELGADO-NOGUERA M.: Factors associated with length of hospital stay in minor and moderate burns at Popayan, Colombia. Analysis of a cohort study. *Burns 42*, 1 (Feb. 2016), 190–195. 2

[ZAS*13] ZOLFAGHAR K., AGARWAL J., SISTLA D., CHIN S.-C., BASU ROY S., VERBIEST N.: *Risk-O-Meter: an intelligent clinical risk calculator*. an intelligent clinical risk calculator. ACM, New York, New York, USA, Aug. 2013. 1

[ZGP15] ZHANG Z., GOTZ D., PERER A.: Iterative cohort analysis and exploration. *Information Visualization 14*, 4 (Sept. 2015), 289–307. 1