

Visual Cluster Analysis in Support of Clinical Decision Intelligence

David Gotz, PhD¹, Jimeng Sun, PhD¹, Nan Cao, MS², Shahram Ebadollahi, PhD¹
¹IBM T.J. Watson Research Center, New York, USA; ²HKUST, Hong Kong, China

Abstract

Electronic health records (EHRs) contain a wealth of information about patients. In addition to providing efficient and accurate records for individual patients, large databases of EHRs contain valuable information about overall patient populations. While statistical insights describing an overall population are beneficial, they are often not specific enough to use as the basis for individualized patient-centric decisions. To address this challenge, we describe an approach based on patient similarity which analyzes an EHR database to extract a cohort of patient records most similar to a specific target patient. Clusters of similar patients are then visualized to allow interactive visual refinement by human experts. Statistics are then extracted from the refined patient clusters and displayed to users. The statistical insights taken from these refined clusters provide personalized guidance for complex decisions. This paper focuses on the cluster refinement stage where an expert user must interactively (a) judge the quality and contents of automatically generated similar patient clusters, and (b) refine the clusters based on his/her expertise. We describe the DICON visualization tool which allows users to interactively view and refine multidimensional similar patient clusters. We also present results from a preliminary evaluation where two medical doctors provided feedback on our approach.

Introduction

Motivated by several perceived advantages and hastened by government regulation, adoption rates for electronic health records (EHRs) are increasing across the globe. The primary use case for an EHR is to digitally capture all medical data for an individual patient and to provide efficient access to the stored data at the point of care. Despite the financial investments in information technology required for the deployment and maintenance of EHR systems, these technologies can provide many important benefits, ranging from the reduction of medication errors, to more timely access to medical records, to improved physician communication with both other providers and patients.¹

While enormously valuable, these benefits to traditional care delivery represent just one aspect of EHR technology. A number of secondary uses for EHRs are being explored which exploit the large collections of electronic data that result from EHR adoption. Such applications include, for example, both clinical research² and data-driven quality measures.³ These applications take advantage of population wide statistical data that can be extracted by examining the EHRs for many patients as a group.

Taking this approach one step further are personalized clinical decision intelligence technologies. For a given target patient, these techniques use data analysis algorithms to dynamically identify cohorts of similar patients from within an institution's EHR database. Based on these personalized cohorts of similar patients, the systems then extract statistical data to drive alerts or provide personalized decision support. For example, similar patient analysis has been shown to be effective at near-term prognostics for physiological data.⁴ Others have used patient similarity for risk assessment.⁵

Along these lines, our lab is building a similarity-based decision intelligence system which provides medical professionals managing complex patients with personalized evidence that is extracted from an institution's EHR database. Our approach is to apply statistical cluster analysis algorithms to EHR data to find clusters (which we call cohorts) of similar patients which are relevant to a target patient. Then, once cohorts have been identified, aggregate historical statistics are extracted and displayed to users as added input to their decision making process. This workflow is illustrated in Figure 1.

One critical challenge in this approach is that the similar patient clusters identified by data analysis algorithms are often difficult to understand semantically. Cluster analysis algorithms group similar patients based on statistical patterns. However, because these patterns are hidden within the complex information space of EHRs, it can be challenging for users to understand the semantic differences between statistically significant clusters. Moreover, the clustering may be imperfect for a given clinical task. However, the ability to understand which patients are in each cluster and to allow user refinement of the cluster definitions based on domain expertise is critical to our approach.

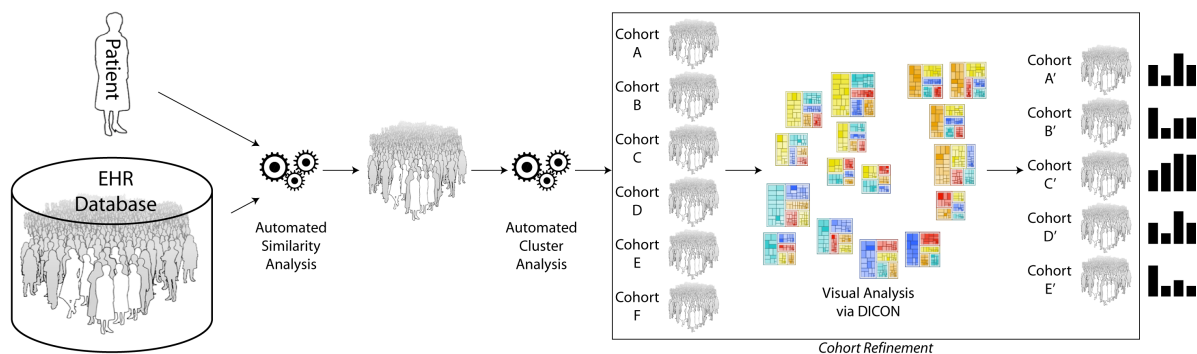


Figure 1: Patient similarity analytics are used to identify a group of EHR records for patients that are similar to a target patient. Cluster analytics are then applied to the set of similar records to produce several different similar patient cohorts. Users can then interactively refine these cohorts based on their expertise using the DICON visual analysis tool described in this paper. Statistics from the clinician-refined cohorts can be used to inform decisions.

To help meet this challenge, we have developed an interactive visualization system which helps domain experts view and refine the similar patient cohorts produced by our analytics. The visualization technique, named DICON, uses treemap-based icons to represent clusters of similar patients. The icons convey multi-dimensional statistical information at a glance and can be manipulated interactively and intuitively to merge, split, and refine the initial clusters into task-appropriate cohorts. These cohorts can then be used as the basis for generating statistical evidence. In this paper we provide an overview of our approach to clinical decision intelligence, describe the DICON visualization which we developed for cluster analysis, and share feedback we received from physicians who were given access to our software.

Background

The secondary use of EHR data is a topic that has received increasing attention as EHR adoption proliferates. Significant attention has been given to improving overall health policy and to developing a framework that would open health data for new applications.^{6,7} Such frameworks would significantly lower the barriers for new technology development and deployment.

Benefits of the broader use of EHR data have been demonstrated in a number of research projects. Most relevant to the work presented in this paper are systems that have analyzed large databases of EHR data to find sets of similar records. Such “patient similarity” approaches have been explored in a variety of practice areas ranging from emergency rooms to risk scoring. For example, Orthuber and Sommer developed a similarity-based search tool for patient records that is used for decision support.⁸ A slightly different approach was adopted by Wongsuphasawat and Shneiderman who used visualization-based techniques to interactively identify similar records.⁹ Both of these techniques help users identify individual similar records which can be used anecdotally to inform decision makers.

Another class of algorithms uses aggregate statistics from clusters of similar patients as an added input when making difficult decisions. For example, Ebadollahi et al. used similar patient records to improve near-term prognosis of physiological data.⁴ For a given patient, their system retrieved a cohort of statistically similar patients and analyzed aggregate statistics from the cohort’s historical physiological data to accurately predict when adverse events were likely to occur. Following a related approach, Chattopadhyay et al. utilize historical data from similar patient records to calculate suicide risk.⁵ While powerful, these techniques rely upon clusters of similar patients which are determined by complex algorithms. As a result, it can be difficult for doctors to understand the characteristics of patients in a cluster. In addition, automatically generated clusters can often require manual adjustment by domain experts yet this capability is typically missing or very limited.

Because of these challenges, which are universal across many application areas that rely on clustering algorithms, several information visualization techniques have been designed for these tasks. These range from scatter plots^{10,11} to parallel coordinates^{12,13} to heat maps.^{14,15,16} These techniques can be highly effective under various conditions.

However, they typically do not scale well for large numbers of clusters and can be difficult for users to follow. Most importantly, these techniques support little or no refinement of the initial clustering structure produced by underlying analysis algorithms. Unfortunately, these limitations are problematic for the clinical applications that are a focus of this paper. We therefore use an iconic treemap-based visualization scheme which provides a compact and intuitive multi-dimensional visual cluster representation that scales easily to large sets of clusters. The resulting visualization also provides clear well-defined visual objects which can be easily selected by users for interactive manipulation at multiple scales.

A final area of information visualization work related to DICON is in the use of icon-based visual representations.^{17,18,19,20} Such approaches are designed to be easily accessible to non-expert users, effective for multi-dimensional data, and easy to manipulate via user interaction. A limitation, however, is that icons are often limited in the amount of information they can convey. DICON embraces many of the benefits of these tools while embedding a large amount of information about both overall cluster statistics and individual entity properties that are often missing in classic icon-based designs.

Clinical Decision Intelligence Using Patient Similarity

Adopting an effective EHR system provides many benefits, such as improved accuracy and information sharing, when used as a straightforward replacement for traditional paper records. However, as described earlier in this paper, the databases of medical information produced by such systems can be exploited in many valuable secondary ways. In particular, as EHR databases grow sufficiently large, they can be mined to extract statistically significant insights about personalized populations of patients.

Along these lines, we are developing a similarity-based clinical decision intelligence system which provides medical professionals responsible for complex patients with personalized evidence extracted from an institution's EHR database. Our approach is to apply similarity and cluster analysis algorithms to EHR data to find clusters of patients which are relevant to a medical decision. This workflow is depicted in Figure 1. For a given patient, similarity analysis produces a set of the most similar EHR records. However, these similar patients are similar in many different ways. For instance, a patient with several co-morbidities might have different groups of patients who are relevant to each of her underlying problems. We apply cluster analysis algorithms to subdivide the overall similar patient cohort into a number of statistically interesting clusters.

While the cohorts produced by cluster analysis can be used directly as the basis for clinical intelligence generation, clinicians often need to explore and refine the cohorts based on their domain expertise. We refer to this stage as *cohort refinement*. Refinement is valuable because cluster analysis algorithms detect statistical patterns, often with little or no *a priori* semantic knowledge. As a result, these automated algorithms can produce cohorts that are hard for clinicians to label semantically. However, semantically meaningful cohorts are required if the statistical insights extracted from the cohorts are to be used clinically.

To enable interactive cohort refinement by domain experts, we have developed a new visualization technique which we call *Dynamic Icons*, or DICON.²¹ Using DICON, clinicians can interactively explore the clusters produced by the automated analysis step and judge their quality. In addition, DICON lets users intuitively manipulate clusters of patients via drag and drop techniques to merge and/or split groups of patients based on domain expertise. We describe DICON in more detail in the next section.

Consider a user who is making a medication order decision for a specific cancer patient. Using DICON, this user can apply his/her domain expertise and contextual knowledge to refine the initial set of algorithmically determined similar patient clusters into cohorts that are more decision-appropriate. After refinement, historical statistics are then extracted for each cohort and presented as supporting data to aid in the user's decision. For example, in our prototype system we present a target patient's lab test results in the context of aggregate lab test results for various similar patient cohorts who have undergone alternative disease-appropriate medication treatments. An example of this display is shown in Figure 2. Following a similar workflow, such an approach is useful not only for clinicians but also for other professionals such as medical directors and researchers.

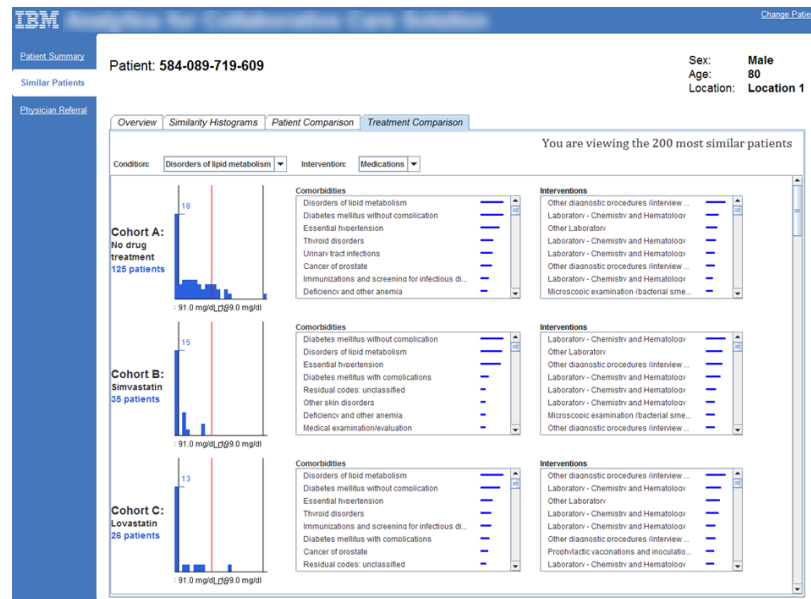


Figure 2: Statistics for each cohort of similar patients are presented using histograms in our web-based prototype system. This view shows how the target patient's lab results in the context of results for various similar patient cohorts.

DICON: Visualization Support for Cluster Analysis

DICON is an interactive visualization tool designed for cluster analysis. It uses dynamic icons to represent clusters of data as shown in Figure 1. In this section we first describe the design principles we followed when developing DICON. We then describe the visual encoding methodology employed by the DICON visualization. Next, we introduce three key user interactions which enable dynamic user-driven cluster manipulation. Finally, we provide a brief overview of the DICON system. A formal description and evaluation of DICON from a visualization perspective is beyond the scope of this paper and is available elsewhere.²¹

Visual Design

While exploring solutions for the problem of cohort refinement, we identified four central design principles that guided the development of DICON. Specifically, we determined that the DICON visualization must provide:

- **Multi-Granularity.** Multidimensional EHR data contains a wide variety of information. An effective design should be able to show various types of information distributions, data variances and diversities at different levels of detail.
- **Consistency.** A visualization design should apply a uniform visual encoding across data types so that users can smoothly switch between different information concepts. In particular, our design utilizes the same set of visual properties and features to represent a range of data from individual patients to patient clusters.
- **Stable Spatial Organization.** Patient features, patients, and patient clusters should be spatially organized such that positions encode meaning. Data updates, such as redefining cluster relationships, should be visually reflected in a stable manner to maintain a user's mental map as much as possible.
- **Rich Interactivity.** A rich set of user interactions should be supported to enable intuitive exploratory analysis and refinement of patient clusters.

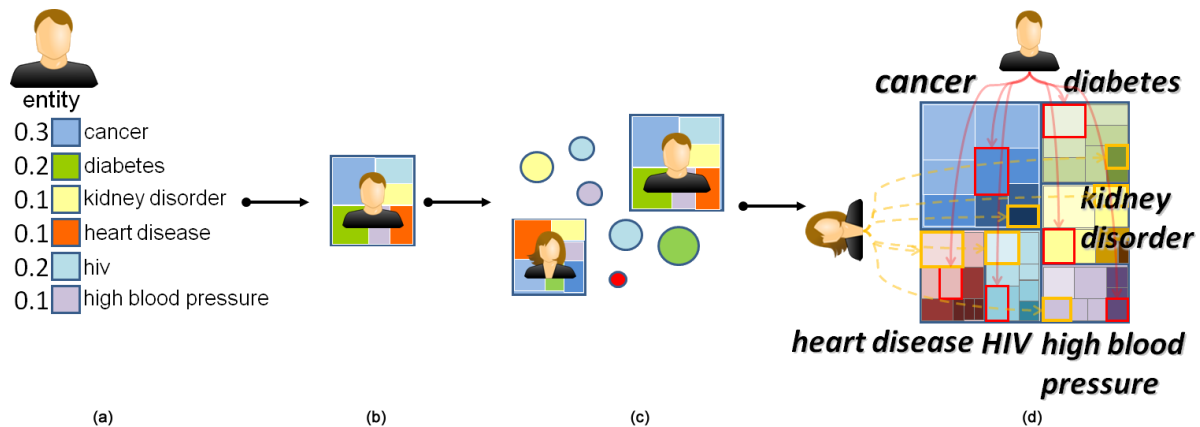


Figure 3: DICON uses an icon design that encodes (a) a feature vector for each patient as (b) a series of color coded regions. This process is (c) repeated for all patients and (d) clusters are represented as interleaved hierarchical arrangements of the color-coded regions. The overall icon conveys the underlying prominence of each dimension of the feature vector across the cluster through the total area allocated to each color.

Visual Encoding

Following the design principles listed above, we designed a Dynamic ICON visualization technique which represents clusters of multidimensional patient data as compact glyphs. The design uses a combination of spatial size, position, color, and opacity to convey key cluster properties. The visual encoding for our design is illustrated in Figure 3.

As shown in Figure 3(a), patients are described by a set of numerical attributes. These values are derived from a patient's EHR. A subset of these attributes are selected as features to be represented in the visualization. DICON visually represents each of these patient features using a colored rectangle. The color of the rectangle indicates the type of feature while the area indicates the feature value. Feature values are normalized to a common scale (e.g., between 0 and 1) to allow the visualization of multiple features in the same icon regardless of scale. The rectangles are packed together to form an iconic representation of the patient as shown in Figure 3(b).

When a cluster contains more than one patient, the individual patient icons must be combined into a single aggregate iconic representation. We generate a cluster's icon by splitting each patient's icon into the individual feature rectangles and repacking these rectangles after grouping them by feature type. This is done using a treemap-based layout where a cluster serves as the top level object, feature types form the second level of the hierarchy, and individual patients make up the third and final level of the hierarchy. The size of a cluster icon represents the total number of entities in that cluster. For examples, the total area for an icon representing a 20 patient cluster will be twice the size of an icon for a 10 patient cluster. We use rectangular treemaps²² as the base structure for our icons and apply the squarified treemap layout algorithm²³ to obtain desirable aspect ratios for the rectangular cells.

Each cell is normally rendered with full opacity, resulting in the same color for all cells for a given feature type. However, color opacity can also be mapped to one of several statistical measures to highlight various cluster properties. For example, if a user wants to see a visual representation of cluster consistency, she can set the color opacity for cells to reflect the difference between a cell's value and the cluster's mean value for the given feature. In this way, outliers can be made to stand out from cells that are close to the mean.

This design brings a few key advantages. First, it compresses high dimensional cluster information within relatively small cluster icons which can be easily embedded within other visualizations. For example, Figure 6 shows the icons embedded within a scatter plot visualization. In addition, the design provides several visual cues that facilitates exploratory analysis. Finally, our design scales well to large numbers of clusters as shown in Figure 4. Yet there are also some limitations to our approach. In particular, the number of feature dimensions that can be visualized at any



Figure 4: A screenshot of DICON visualizing 50 clusters, illustrating DICON's ability to handle large numbers of clusters.

one time is limited because each must be represented by a unique user-distinguishable color. To alleviate the problem, feature selection can be used to identify the key features that should be included in a visualization.

User Interaction

DICON provides a number of dynamic interactions that can be used by users to refine patient clusters:

Split. Given a cluster icon, users can drag one or more patients out of a patient cluster. This user-driven cluster enhancement action results in splitting the original cluster into two parts. We animate the transition during a split so that users can visually follow the change in groupings.

DICON also provides an intelligent split interaction which users can initiate by right clicking on a cluster. This provides a context menu from which users can choose a specific cluster division algorithms to apply. DICON supports both binary split and outlier split algorithms. The binary split option splits a cluster into two even clusters. The outlier split moves the 10% of patients with the largest deviations from the cluster mean to a new separate cluster.

Finally, DICON allows users to split clusters by fixed criteria along certain metadata properties. For example, users can cluster patients by age, sex, or location.

Merge. The inverse of split, users can merge two clusters together by dragging the icon for one cluster and dropping it on the representation of a second cluster. In addition, users can drag a selection lasso around a group of 2 or more clusters to have them all merged together into a single cluster.

Filtering. The filtering interaction allows users to turn on or off different feature types such as cancer and diabetes shown in each cluster. When a feature type is filtered, the corresponding visual elements are hidden and the icons are repacked. This interaction helps users to drill down into a subset of features that are most relevant for a given analysis.

System Overview

The DICON architecture, shown in Figure 5, consists of three primary components. First, a *preprocessing module* extracts key features from a multidimensional dataset and conducts a cluster analysis based on these features. In our clinical application, this is the portion of the system which automatically generates an initial set of similar patient cohorts. The *visualization module* maps the patient features in each cluster to a multivariate visual display according to the visual design described above. It employs custom algorithms for laying out clusters of entities by considering their relations in multiple granularities. It also includes pattern enhancement capabilities that improve the overall

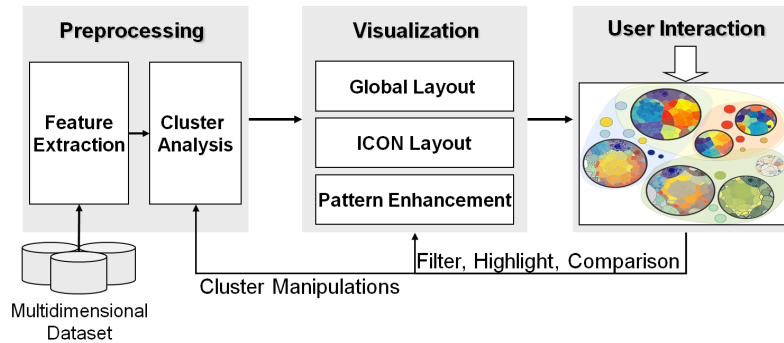


Figure 5: The DICON visual analysis system.

appearance and legibility of the visualization. The *user interaction* manages the interactive features described above, allowing users to explore the cluster results and adjust them efficiently. These operations feed back into the preprocessing and visualization modules to enable user-driven data exploration. The implementation of this system is a web-based application that uses a combination of Java, HTML, JavaScript and JSP technologies.

Global Layout A key responsibility for the visualization module is global layout. When a set of icons are generated, various layout algorithms can be used to lay them out spatially across a visualization canvas. For example, when the icons are used to represent geographical patient clusters, they can be laid out based on their physical locations. The icons can also be embedded within more abstract spaces such as scatter plots (see Figure 6) or timelines.

DICON also provides a MDS-based projection to layout cluster icons based on their similarity. Furthermore, to avoid overlap, a fast overlap removal algorithm²⁴ is adopted. It removes all overlaps while retaining each icon's original position as much as possible. Some improvements were made to these algorithms to facilitate interactive cluster manipulations. First, we minimize the movements when cluster changes occur by smoothing positional changes based on objects' previous positions. Second, an incremental layout technique is used in support of split and merge commands. For example, when patients are split off from a cluster, only the modified cluster and the newly split patients are re-laid out in a sub-area followed by a global overlap removal. In this way, the positions of other cluster icons not impacted by the split operation do not change.

Results: Clinical Application and Physician Evaluation

To begin evaluating the DICON tool's ability to support cohort refinement as outlined in Figure 1, we performed a case study where we asked two physicians to provide feedback on our prototype system. Both subjects in the case study are former emergency room physicians with several years of clinical experience. In addition, the two subjects have held managerial roles which give them an appreciation of how management staff (e.g., medical directors) would use such tools.

To gather feedback on our approach, we spent 30 minutes with each participant. After a few minutes spent reviewing the visual design and user interactions that DICON supports, the remainder of each session we spent refining patient cohorts and discussing various aspects of the visualization environment. The session moderator posed questions to the users and recorded notes throughout the experiment to capture the physicians' feedback.

During the instructional portion of the evaluation, many questions were asked about the visual representation. The fact that each icon represented a single cluster was immediately clear. However, the treemap-based interior structure for each icon required significantly more explanation. In particular, both physicians took some time to understand how the hierarchical arrangement of cells distributes the features of a single patient spatially across an icon. One physician felt that the representation was "complex," especially when looking at individual patients. However, once the design concept was fully explained, users were able to see at a glance several properties of each cluster.

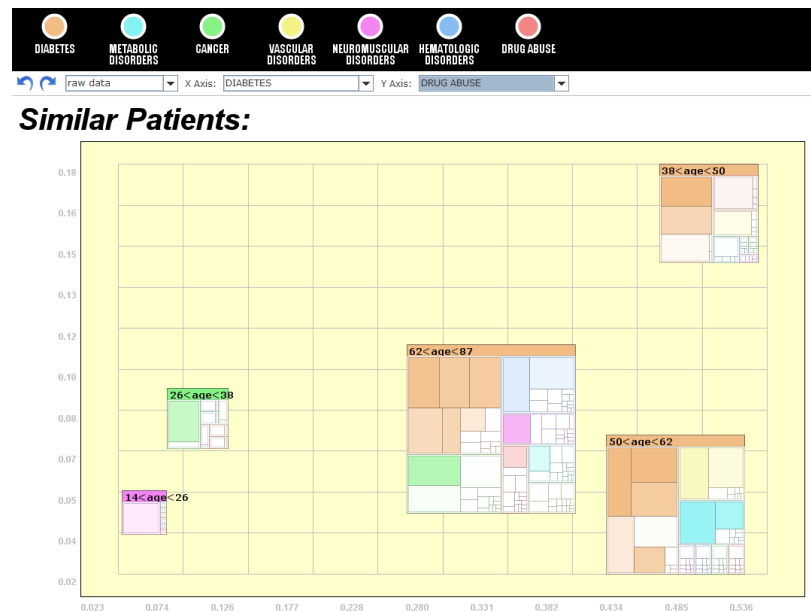


Figure 6: Icons can be embedded within other visualizations such as scatter plots. This screenshot shows that diabetes (x-axis) becomes common in older patients. Meanwhile, drug abuse (y-axis) is most frequent within only the 38-50 age bracket although that cluster's icon shows that diabetes is still a much larger concern.

When introducing participants to DICON, the illustration in Figure 3 was especially useful in helping to convey how the icons were constructed. In addition, an interactive demonstration of the tool was extremely valuable because the animated transitions when splitting a single outlier patient off from a cluster clearly highlighted how various features for the patient were located throughout the icon. Upon seeing the animation, one participant exclaimed, “Oh, I get it! That makes a lot of sense.”

As expected, our choice to embed detailed information about each cluster into the icon caused a significant increase in complexity. This is certainly a drawback of our approach. However, we feel that the benefit of adding the added information to our icons far outweighs the cost in visual complexity because without the additional information users would not have access to data needed to perform cluster refinement. Moreover, users can ignore the interior structure of our icons to gain a high-level overview of cluster properties without the added complexity.

Overall, the physicians were both intrigued by the DICON visualization and felt that it provided value. “It provides an interesting way to define cohorts” said one physician, who was especially interested in the drag-and-drop nature of the technique. He felt that DICON provided a “very intuitive interface” for manipulating sets and very much liked the icon design which provided a concrete object for him analyze and manipulate. When referring to the interactive refinement of cohorts, one physician stated that “as a medical director, this is exactly what I would want to do.” The icon design let him “do it rapidly [via] drag and drop” instead of “giving it to a programmer” to generate a new report.

In addition to commenting on DICON's current functionality, the participants also made suggestions for future improvements. For example, one user wanted to have more powerful rule-based filtering capabilities. While the tool does allow you to re-cluster according to individual dimensions (e.g., re-group clusters by age), the physician wanted the ability to do this for combinations of dimensions (sex and age). This is a feature that we hope to introduce in future revisions of the tool.

A more complicated request made by one user was the ability to drag the icon for a cohort from our tool onto icons for other system functionality. His suggestion was to use this approach to issue requests for additional analytics to be applied to a given group of patients. The user's request for this feature shows that the tangible icons we designed for representing cohorts form a very powerful representation in the minds of our users. The icon itself has become the

object that the user wishes to operating on. We believe this is a very powerful design approach and we are exploring ways to adopt it.

Conclusion

This paper described a similarity-based clinical decision intelligence system which provides users with personalized evidence that is extracted from an institution's EHR database. We apply statistical cluster analysis algorithms to EHR data to find clusters of patients who are relevant to a clinician's target patient. Then, based on aggregate statistics extracted from these clusters, we provide personalized decision intelligence to clinicians as an added input to their decision making process.

A critical component in this system is a visualization tool—DICON—that allows clinicians to understand and refine patient cohorts interactively. DICON uses treemap-based icons to represent clusters of similar patients. The icons convey multi-dimensional statistical information at a glance and can be manipulated interactively and intuitively to merge, split, and refine the initial algorithm-generated clusters into expert-defined task-appropriate cohorts. The paper provided an overview of the visual design of DICON and described its interactive features. The initial feedback received from physicians shows that the visualization is accessible to users without significant training. Moreover, the direct manipulation made possible by the visualization's interaction capabilities is attractive for the cohort refinement task we aim to support.

While the prototype implementation described in this paper shows promise, there remain several topics for future work. First, the results from our initial evaluation must be further validated via larger, more rigorous users studies. The feedback from any future studies will certainly motivate design improvements in our visualization as the system evolves. In addition, we hope to make progress on many of the valuable suggestions made by the physicians in our initial case study. For example, the suggestion to use cohort icons as tangible objects that users can drag and drop onto other system components may be a very useful extension to our current application.

References

1. Catherine M. DesRoches, Eric G. Campbell, Sowmya R. Rao, Karen Donelan, Timothy G. Ferris, Ashish Jha, Rainu Kaushal, Douglas E. Levy, Sara Rosenbaum, Alexandra E. Shields, and David Blumenthal. Electronic health records in ambulatory care a national survey of physicians. *New England Journal of Medicine*, 359(1):50–60, 2008.
2. John Powell and Iain Buchan. Electronic health records should support clinical research. *Journal of Medical Internet Research*, 7(1), 2005.
3. Paul C Tang, Mary Ralston, Michelle Fernandez Arrigotti, Lubna Qureshi, and Justin Graham. Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: Implications for performance measures. *Journal of the American Medical Informatics Association*, 14(1):10–15, January 2007.
4. Shahram Ebadollahi, Jimeng Sun, David Gotz, Jianying Hu, Daby Sow, and Chalapathy Neti. Predicting patient's trajectory of physiological data using temporal trends in similar patients: A system for Near-Term prognostics. In *Proceedings of the American Medical Informatics Association Annual Symposium (AMIA)*, Washington, DC, 2010.
5. S. Chattopadhyay, P. Ray, H. S Chen, M. B Lee, and H. C Chiang. Suicidal risk evaluation using a Similarity-Based classifier. In *Proceedings of the 4th international conference on Advanced Data Mining and Applications, ADMA '08*, page 5161, Berlin, Heidelberg, 2008. Springer-Verlag.
6. Charles Safran, Meryl Bloomrosen, WEdward Hammond, Steven Labkoff, Suzanne Markel-Fox, Paul C Tang, Don E Detmer, and With input from the expert panel (see Appendix A). Toward a national framework for the secondary use of health data: An american medical informatics association white paper. *Journal of the American Medical Informatics Association*, 14(1):1–9, January 2007.

7. Meryl Bloomrosen and Don Detmer. Advancing the framework: Use of health DataA report of a working conference of the american medical informatics association. *Journal of the American Medical Informatics Association*, 15(6):715–722, November 2008.
8. Wolfgang Orthuber and Thorsten Sommer. A searchable patient record database for decision support. *Studies in Health Technology and Informatics*, 150:584–588, 2009.
9. Krist Wongsuphasawat and Ben Shneiderman. Finding comparable temporal categorical records: A similarity measure with an interactive visualization. In *IEEE Visual Analytics Science and Technology*, 2009.
10. Edward R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, February 1992.
11. Daniel B Carr, Richard J Littlefield, and Wesley L Nicholson. Scatterplot matrix techniques for large n. In *Proceedings of the Seventeenth Symposium on the interface of computer sciences and statistics on Computer science and statistics*, page 297306, New York, NY, USA, 1986. Elsevier North-Holland, Inc.
12. Alfred Inselberg and Bernard Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proceedings of the 1st conference on Visualization '90, VIS '90*, page 361378, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.
13. Matej Novotny. Visually effective information visualization of large data. In *8th Central European Seminar on Computer Graphics (CESCG 2004)*, pages 41–48, 2004.
14. Sharlee Climer and Weixiong Zhang. Rearrangement clustering: Pitfalls, remedies, and applications. *The Journal of Machine Learning Research*, 7:919943, December 2006.
15. M B Eisen, P T Spellman, P O Brown, and D Botstein. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95(25):14863–14868, December 1998.
16. M Friendly. Corrgrams: Exploratory displays for correlation matrices. *The American Statistician*, 56:316–324, 2002.
17. RM Pickett and GG Grinstein. Iconographic displays for visualizing multidimensional data. In *Systems, Man, and Cybernetics, 1988. Proceedings of the 1988 IEEE International Conference on*, volume 1, pages 514–519, 1988.
18. Frits H Post, Frank J Post, Theo Van Walsum, and Deborah Silver. Iconic techniques for feature visualization. In *Proceedings of the 6th conference on Visualization '95, VIS '95*, page 288, Washington, DC, USA, 1995. IEEE Computer Society.
19. Herman Chernoff. The use of faces to represent points in K-Dimensional space graphically. *Journal of the American Statistical Association*, 68(342):361–368, June 1973.
20. Daniel A. Keim and Hans-Peter Krigel. VisDB: database exploration using multidimensional visualization. *IEEE Computer Graphics and Applications*, 14(5):40–49, 1994.
21. Nan Cao, David Gotz, Jimeng Sun, and Huamin Qu. DICON: interactive visual analysis of multidimensional clusters. In *Proceedings of the IEEE Information Visualization 2011, InfoVis 2011*. IEEE Computer Society Press, 2011.
22. B. Shneiderman. Tree visualization with treemaps: 2-d space-filling approach. *ACM Transactions on graphics (TOG)*, 11(1):92–99, 1992.
23. M. Bruls, K. Huizing, and J. van Wijk. Squarified Treemaps. In *In Proceedings of the Joint Eurographics and IEEE TCVG Symposium on Visualization*. IEEE, 1999.
24. T. Dwyer, K. Marriott, and P. Stuckey. Fast node overlap removal. In *Graph Drawing*, pages 153–164, 2006.