Report

UNIVERSITY OF SKÖVDE

VISUALIZATION IN HEALTH

Visual Data Analysis, IT740A

Group 4

Welemhret Welay Baraki

a20welba@student.his.se

Submission date: 2021-03-04

Contents

Abstract

In this era of technology handling accumulated data in different repositories to give organized services for decision making and inferring patterns is challenging especially in the health sector. So effective analysis and visualization techniques are required to extract valuable insights to help professionals and patients. In this review, we assess and analyse four articles focused on health care data analysis and visualization. The first article tried to address the psychological and social aspects of patients, and the second article deals on creating patterns from health care big data, the third article tried to partition the tasks of big data analytics and visualization into different layers and the last article discusses on the detailed visualization and modelling of disease progression pathways using HMM.

1. Introduction

High volumes of big data[1][2] are accumulated from different sources and in different formats. These high volumes of data are available in different sectors one of them is the health sector. Extracting knowledge and insights from the raw data with the interference of humans is very difficult. Additionally, the extracted knowledge and insights (analysis) should have to deliver to the professionals convincingly and persuasively that is the art of visualization. Accordingly, health professionals will be able to get and extract insights, patterns, and knowledge in a very persuading and organized manner.

The main theme of the selected papers deals with the health data how patients have effectively cured with the help of AI-Powered analysis with narrative story-telling to incorporate the psychological and social aspects[3] with martini glass structure, analyzing health care data to mine patterns and associations of different of data(e.g. diseases)[2], partitioning health data analysis and visualization into manageable distinct layers[4] which introduces ease and flexibility and simulating and visualizing disease progression pathways[5].

The main contribution of these visualization projects and researches will leap forward to the health sector in multiple dimensions. The articles generally contributed on different aspects of the health sector are listed as follows:

- By extracting emotions and feelings from the social media data and socio-economic data, which is the neglected part of the medical treatment gives awareness and focus and that will be a valuable input to the health professionals[3].
- Demonstrating a method for visualizing smart data from health data, as well as associated information and expertise, which facilitates healthcare analytics[2].
- A mechanism was built on enabling the composition of analysis and visualization components in a customizable workflow[4].
- Creating disease progression models, summarizing disease states, interactively exploring disease progression processes, and constructing, assessing, and comparing clinically important patient subgroups[5].

Generally, these articles contribute a lot to the health sector of visual data analysis on quantifying and communicating the difficult parts of psychological and social aspects[3], demonstrating how health care big data[2] is handled and analyzed, developing customizable workflow[4], and visualizing disease progression[5].

2. Visualization in Health

In the following sections, we will discuss the selected visual data analysis articles of the health sector focused on medical treatment, disease progression pathways, managing healthcare data for effective services, and component-oriented visual tools that handle different varieties of health care data.

2.1 Humane Visual AI: Telling the Stories Behind a Medical Condition

The main theme of this article is on visualization in the health sector with narrative storytelling, specifically on medical treatment of patients by considering their psychological and social backgrounds for effective treatment with respect to their biological aspect which is basically not considered in the traditional biomedical model. The researchers basically identify the psychological and social aspects from different kinds of literature and discussion on the works that use dashboards which that they identify problem of information overload to render multidimensional data in a single screen.

This work mainly emphasizes undertaking to quantify and visualize the psychological and social aspects of patients. Throughout the development of the project for the analysis and quantification of the psychological and social aspects they use a deep learning framework to extract medical entities from the Reddit posts, Dictionary-based to extract emotions from the posts using LIWC[1] dictionary category 'body' and Wikipedia 'human body and organs', and Probabilistic approach to compute the prevalence of drug prescription related to the condition. Additionally, for the visualization of these aspects, layered Martini Glass visualization was used.

This article tried to combine the biological, psychological, and socio-economic[3] aspects of patients to develop a persuasive visualization tool to help doctors for the effective diagnosis and treatment of patients. The biological aspect is handled by capturing the symptoms/drugs association extracting from different health-related websites and developed a Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) database. Even though the psychological and socio-economic aspects are hard to quantify, the researchers studied data collected from Reddit(psycho) and census data(socio-economic) in England. They developed deep learning to extract the psychological aspects of users and a probabilistic model to estimate the prevalence of drug prescriptions with census data. Finally, visually communicate each medical condition through narrative-style layered Martini Glass visualization that created a story based on the analysis of the three aspects, reduce the complexity of the bio-psycho-social model. In the visualization of the Bio-Psycho-Socio model bubbles are used for the biological, human body with heat-texture and emotions for the psychological, and a choropleth map with radar chart for the social is used by the researchers.

The most important aspect that is addressed in this model is that incorporating the psychological and social aspects of patients integrated with the biological medication will transform the next generation of health care systems. The main problem that they face during the social media extraction is that extracting patterns from unstructured data is difficult. To identify the visualization design goals they prepared open-ended questions and asked health experts on intuitiveness, the interest of the general public, and additional features. Including medical experts in the research part of the elicitation study is

[1] Stands for "Linguistic Inquiry and Word Count"

good, but the extraction of the emotions (psychological aspects/conditions) from the social media are not involved to address the feelings/disease associations.

User studies with 108 participants are performed to test questions that discuss the degree of adoption of the visualization paradigm and the baseline. Generally, users have answered personal and general questions in AI and the three aspects of health care. The results of the user study suggested that developing an intelligent visualization tool is important for creating awareness and supporting effective medication of patients by incorporating the psychological and social aspects of patients. The observation is limited to 14 conditions which are not exhaustive can be considered as a limitation in this research. The most compelling graphic elements described by many users are body-part simulation and bubbles of symptoms and diseases. Their response to socio-economic considerations is relatively small. Additionally, Biases in social media data and testing one visual storytelling instead of including others are considered as a weakness of the established model.

Overall, the convincing influence of the visualization was greater than that of the baseline in which users provided a favourable answer to the model. This visualization research indicates that strengthening health communication through AI visual analysis and integrating the model into social media channels would allow the comprehension of the patient's bio-psycho-socio level.

2.2 Big Data Analysis and Services: Visualization of Smart Data to Support Healthcare Analytics

The main theme of this article is how analysis and visualization of accumulated data of the health sector are vital for effective healthcare services by determining the associated frequent patterns using the HSL (Hue-Saturation-Lightness) visualization tool with pattern mining algorithms. The core task conducted in the analysis part of this article was to extract patterns using the pattern mining algorithms to identify the associations and frequency of disease from the accumulated health care data and they visualized using HSL visualization. The project aims to reveal the visibility of health care data to users to be able to extract knowledge and insight using pattern mining algorithms.

The visualization model uses a circle structure to illustrate frequent patterns and their relationship—both prefix-extension relationships and subset-superset relationships—uses Color to reflect the cardinality of frequent patterns. Instead of using arbitrary color with different cardinalities, the tool makes effective use of the hue ranges from 0° to 360° in the HSL color model to find a systemic way to determine the color, to visualize all frequent patterns, the designed the visualization to put them in a radial layout. Based on the number of domain items and their extensions, HSLviz splits the innermost circle into several sectors. The main advantage of using this model is that it provides the users with freedom for exploration by supporting details on subset and supersets separately.

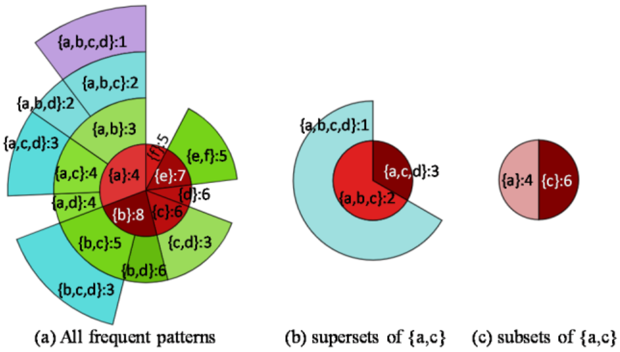

Figure 1 HSLviz showing (a) all frequent patterns, (b) supersets of {a, c} (c) subsets of {a,c} [2]

Additionally, the tool has the capability of ordering the frequent patterns based on their alphabet and frequency. The visualization model was evaluated for two healthcare datasets (*US National Health Interview Survey* and *Surveillance, Epidemiology and End Results* (*SEER*) *Program dataset)* associated with cancer patients by running frequent pattern mining algorithms on the datasets. The evaluation method is not clearly stated in the paper. They simply put what they observe from the experiments of the two data sets on the analysis and visualization results. Overall, the visualization tool is good but their evaluation is not clear especially on the persuasiveness and acceptance level of the visualization to users. The advantage of the new model is its flexibility, includes many different additional features, and shows the connection between related items.

2.3 HealthVision: a web-based platform for healthcare data analysis and visualization

The main theme of this article is to visualize different varieties of health care data from different sources using different machine learning and deep learning analysis algorithms and plenty of visualization components in a custom application-specific data flow based on the component-oriented design methodology which gives them flexibility for their development.

HealthVision[4] is a workflow-based architecture that partitions workflows into four layers: Raw Data accepts and processes data from different sources and in different formats, Data Analytics which used machine learning algorithms to analyze the data accepted from the Raw Data layer, Decision Support layer mainly uses a rule-based approach to select the best visualization method for the data and finally the data visualization layer presents the data in different visualization components based on the recommendation of the decision support layer. Throughout the design and implementation of the project ReactiveX, AngularJS, and python programming language were used. Variety of data analysis machine learning algorithms are available in their design to use, and different variations of Line-based plots, Bar-based plots, point plots, Heatmaps, Graph-based visualizations, and spatial maps are used to visualize the healthcare data. Partitioning into layers created the opportunity to control each layer independently and the freedom to select different approaches in different components for a variety of data formats.

The researchers demonstrate[4] their visualization tool using different types of data from different sources of healthcare data for clustering, visualization, clinical analysis based on similarities such as Sjogren's syndrome (SS), Parkinson's disease, chronic lymphocytic leukemia, asthma, and frailty.
Even though their user evaluation results are not presented, in the paper, they stated that they got a promising result in data exploration and event detection from the preliminary usability test with patients and clinicians.
Finally, evaluation of the platform in other types of clinical data to support large-scale studies, incorporation of deep neural network methods, decision support techniques based on machine learning algorithms, and progressive visual analytics techniques are forwarded as a future research direction for researchers in the field.

2.4 DPVis: Visual Analytics with Hidden Markov Models for Disease Progression Pathways

The main focus of this article is on visualizing disease progression pathways using the Hidden Markov Models and different visualization techniques to be able to ease understanding of different disease progression pathways and help health professionals for effective treatment of patients accordingly. The target users were clinical researchers a series of discussions and conferences are taking place to identify the goals and needs of the health professional in the area. Then, the researchers consolidate research questions and drive hypotheses on optimal treatments of patients by observing their disease progression patterns, discovering distinct states of patients, and detecting pre-symptomatic progression signals. Three main goals are identified to achieve in their visualization tool. These goals are visualizing disease progression to be able to identify distinct disease progression states, discovering heterogeneous trajectory groups, and find associations between scientific trajectories and variables. The data collected includes multiple visits spread over different time points.

Hidden Markov models (HMM) were used to analyze the data collected from the birth cohorts and cross-validation to evaluate the model. The model was tested on type-1 diabetes disease progression patterns of 559 subjects from birth cohort studies using the 11 state HHM model to be able to show progression pathways based on the trajectory groups and autoantibodies. The visualization tool mainly incorporates future matrix, future distribution, state transition patterns, state sequence finder and, creates and refine subgroup lists. Some of the visualization techniques that are used in their research article are heatmap, histograms, beeswarm plots, Sankey diagram, and others are adopted in a convenient way to the domain problem. The strong point of this research is that the data is observed in different ways and using the different visualization features to enable clinical researchers to ease knowledge and insight extraction. Besides, the early involvement of domain experts having open-ended discussions and paired analytics gives advantage of the researchers to have a clear understanding of the problem domain. The visual tool's main feature is creating associations, and identify differences and similarities between multiple subjects which gives insights to clinical researchers. Subgroup builder and state sequence finder gives flexibility to the clinical researchers to be able to view, refine queries, build queries and compare subgroups easily.

The researchers conducted interviews with nine clinical researchers to share their user experiences. Generally, the user’s response was positive in reducing the effort required to analyze observational data, helping users understand HMMs transparently, and multiple views of the visualization help users summarize, search patterns, and build subgroups of disease progression trajectories. Additionally, the problems and limitations that they experience are State sequence Jumps are not handled, Subjects may represent an endotype that are not captured yet, Interpretations of HMM model result is not intuitive, Uncertainties in state assignments, Data collection matters for outcome interpretation, Unknown model configurations, consumes time to learn and interpret, lack of automatic detection of outliers in the pathway and unable to modify states to formulate new states.

Based on the user experience the visualization tool is recommended to consider the knowledge gaps of different professionals from different sectors like statisticians, visualization experts, and health professionals that are involved in. The problems of misinterpretation of the result and losing trust in the visualization were alleviated with series of workshops and discussions.

Finally, they constructed a visual analytics program that allows users to explore disease progression pathways and their interactions with different interventions by combining HMMs and providing views and interactive features to help users propose and evaluate hypothesis.

3. Conclusions and discussion

In [3] visualization of narrative storytelling, with a focus on the medical care of patients and taking into account their psychological and social contexts to include successful treatment with regard to the biological component, which is largely ignored in the conventional biomedical paradigm. The strengths of [3] are quantifying and visualizing the psychological and socio-economic aspects which are difficult to model, simplifying the complexity of visuals using the Martini Glass structure. The limitations of [3] are social media biasness and difficulty of extracting emotions and feelings, health professionals are not included at the early stage of the project, and observations are limited to 14 conditions which are not exhaustive.

The key focus of [2] is how identifying the associated frequent disease patterns using the pattern mining algorithms is critical for efficient healthcare facilities by analyzing and visualizing big data from the health sector with visualization tool of HSL (Hue-Saturation-Lightness). The major strength of this visualization tool[2] that it allows users to explore more freely. The limitation is that the evaluation strategy is not clearly stated especially on the acceptance level of the visualization to users.

In [4] the researchers tried to simulate various types of health care data from various sources using different machine learning and deep learning analysis techniques, as well as numerous visualization components in a personalized application-specific data flow based on the component-oriented, which allows for implementation and control flexibility of different components. To make it easier to consider different disease progression pathways, researchers used the Hidden Markov Model and various visualization techniques in [5]. The limitations are understanding the model is difficult and complex at first use. The strength is easing comparison and creating associations of different groups. The lesson that we should have to take from these articles is that analysis and visualization in the health sector are almost endless and is untapped which requires extensive researches for the betterment of society.

References

[1] P. K. D. Pramanik, S. Pal, and M. Mukhopadhyay, “Healthcare Big Data,” no. October, pp. 72–100, 2018, doi: 10.4018/978-1-5225-7071-4.ch004.

[2] C. Leung, Y. Zhang, C. S. H. Hoi, J. Souza, and B. H. Wodi, “Big data analysis and services: Visualization on smart data to support healthcare analytics,” *Proc. - 2019 IEEE Int. Congr. Cybermatics 12th IEEE Int. Conf. Internet Things, 15th IEEE Int. Conf. Green Comput. Commun. 12th IEEE Int. Conf. Cyber, Phys. So*, pp. 1261–1268, 2019, doi: 10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00212.

[3] W. So, E. P. Bogucka, S. Scepanovic, S. Joglekar, K. Zhou, and D. Quercia, “Humane visual AI: Telling the stories behind a medical condition,” *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 678–688, 2021, doi: 10.1109/TVCG.2020.3030391.

[4] E. Polychronidou, I. Kalamaras, K. Votis, and D. Tzovaras, “Health Vision: An interactive web based platform for healthcare data analysis and visualisation,” *2019 IEEE Conf. Comput. Intell. Bioinforma. Comput. Biol. CIBCB 2019*, 2019, doi: 10.1109/CIBCB.2019.8791462.

[5] B. C. Kwon *et al.*, “DPVis: Visual Analytics with Hidden Markov Models for Disease Progression Pathways,” *IEEE Trans. Vis. Comput. Graph.*, pp. 1–15, 2020, doi: 10.1109/TVCG.2020.2985689.