Visualizing Algorithmic Cause of Death Predictions

Aditya Anerao[1], Maggie Dorr[2] , Chethan Jujjavarapu[2], Andrew Teng[2]

1. Department of Chemistry, University of Washington, Seattle WA
2. Biomedical Informatics and Medical Education Department, University of Washington, Seattle WA

Abstract

Predicting the cause of death is an important problem area in many countries. As a result, algorithms have been developed to predict these causes based on medical information that includes patients' symptoms. We designed visualizations to assist in understanding these cause-symptom relationships and algorithm performances.

Introduction

Verbal autopsy (VA) is an important tool to measure cause of death in populations without complete medical certification of causes of death. This is a common situation in many countries and as a result, building algorithms that can predict the cause of death associated to a VA from medical information is essential to improve healthcare. However, there is copious amounts of data indicating relationships between cause of death and symptoms and different algorithms can predict different outcomes for a single patient. To understand these relationships and better compare algorithms, we have designed the visualizations presented here.

Methods

The data is organized into three outcome groupings: 34, 46, and 55 causes of death. Within each group, there are two types of data:

(1) symptom-causes that indicate which symptoms are related to causes for each patient
(2) probabilities for each cause determined by each specific algorithm.

For data type 1, we calculated the frequency between a cause and symptom and divided by the total number of relationships. These aggregated relationships are presented in Figure 1, 2, and 3. For data type 2, we averaged probabilities for each cause for each algorithm and present in Figure 4.

Visualizations

Figure 1: Force Directed Network

To understand the relationships between causes and symptoms, we start off with a network. The network map allows users to quickly understand the relationships between causes and symptoms at a high-level. True causes and symptoms are distinctly colored, which allows the user to quickly observe the number of causes and symptoms present for each grouping of data. The links encode the strength of the relationship by line thickness. A user is able to click on a desired node to identify the cause or symptom as well as highlight the network and relations between the selected node and the rest of the network.

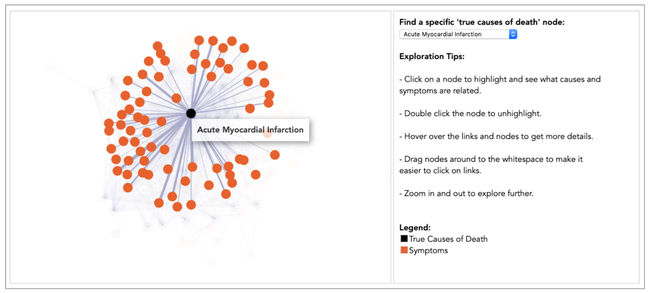

Figure 1: Force directed network

Figure 2: Parallel Coordinate

While the network is informative, it lacks structure to easily understand the data. To address this issue, we developed a parallel coordinate. The parallel coordinate diagram shows the different relationships between causes and symptoms. It allows for more granular observations compared to the network approach. A user is able to interact with the text and lines to highlight specific relationships. For example, we found that drinking is linked to fires (cause of death). In addition, you can select the line to observe the strength of the relationship.

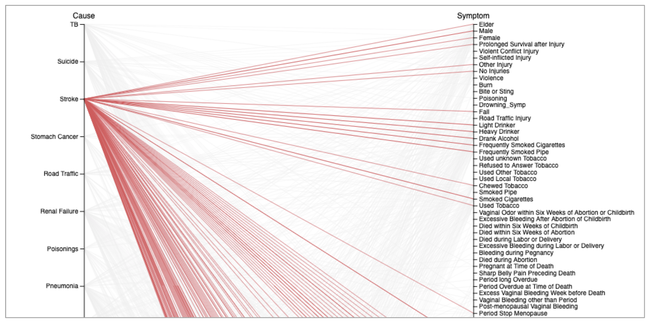

Figure 2: Parallel coordinate

Figure 3: Heatmap

Though the parallel coordinates graph above was more clear in the relationships between causes of death and symptoms, it does not as clearly display the uncertainty in some relationships. Our heatmap delves deeper into the strength of relationship between causes of death and symptoms. Users can easily observe each cause and symptom along the y and x axes, respectively, and at their intersection, view the brightness (i.e., lightness) of the red hue as the level of association. The user can also view the precise value by hovering over the square due to the interactive tooltip feature.

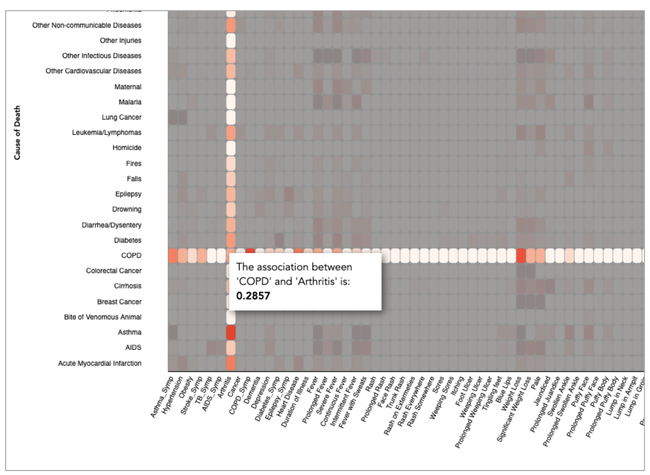

Figure 3: Heatmap

Check out our interactive visualizations by visiting: https://cse512-19s.github.io/FP-Visualizing-Algorithmic-Cause-of-Death-Predictions/

Figure 4: Bar Chart

This graph was designed to compare how different algorithms predict cause of death; the triple bar chart allows all three algorithms to be simultaneously compared. This visualization shows each algorithm's average probability for predicting the correct cause of death. The three algorithms are color coded. This figure serves as an indirect comparison of the performance of different algorithms. Therefore, the height of different bars can be compared for an individual cause of death.

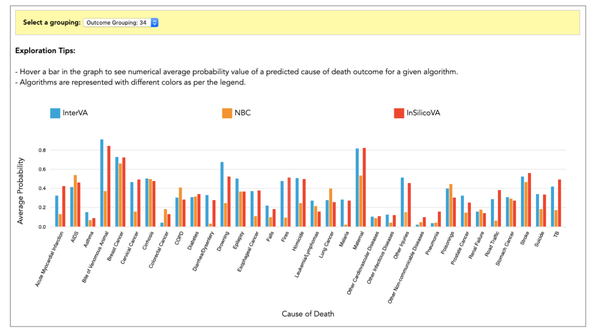

Figure 4: Bar chart

Conclusion

These visualizations allow researchers to easily explore this rich data to observe interesting relationships between causes and symptoms. In addition, the classification performance across algorithms can be compared in order to identify common patterns.

Acknowledgements

We are thankful to our collaborators Dr. Tyler H. McCormick from the University of Washington Department of Sociology and Department of Statistics, and Dr. Zehang Richard Li from the Yale University Department of Biostatistics, whose research generated the datasets and whose guidance inspired this visualization project. We also thank the teaching staff of CSE 512 at the University of Washington, led by Dr. Jeffrey Heer for guidance throughout this project and the course.