Isha Mahadalkar
6th September 2022

THALIS: HUMAN-MACHINE ANALYSIS OF LONGITUDINAL SYMPTOMS IN CANCER THERAPY

Junxiu Tang
Yuhua Zhou
Tan Tang
Di Weng
Boyang Xie
Lingyun Yu
Huaqiang Zhang
Yingcai Wu

[Paper link: https://ieeexplore-ieee-org.ezproxy.lib.purdue.edu/document/9555227]

The paper I decided to summarize this week is named "THALIS: Human-Machine Analysis of Longitudinal Symptoms in Cancer Therapy". It talks about an approach using unsupervised machine learning algorithms to acquire more intelligent data for new cancer patients and survivors based on patients with similar previous diagnoses and symptoms. Their reasoning behind their approach is that cancer patients suffer a myriad of residual symptoms which can be permanent in nature even after going through intense oncologic therapy. These long-lasting symptoms differ in their severity, rate of development, and eventual resolution even amongst patients with the same diagnoses based on a variety of factors. Even though there is a lot of data in the form of questionaries and health records collected from patients, but its difficult to extract any meaningful information due to wide range of symptoms, their partial co-occurrence, unpredictability among patients and across time, location of the tumors and the therapy prescribed.

Their tool uses symptom clusters to understand the underlying mechanisms that drive these clusters and factor-based analysis to estimate the longitudinal symptom evolution of an individual patient. And it provides a vital resource in understanding the nature of the symptoms on a scale wider than just human or machine intelligence and helps in both clinical and symptom research. This tool aims to provide a way for physicians to actually use this data to make informed decisions about their current patients.

One of the main contributions of the work is the actual tool, THALIS: THerapy Analysis of LongItudinal Symptoms, which is a "visual symptom explorer" which runs the algorithms to provide a visual, easy to understand analysis of prior symptom clusters. It shows this data using interactive association-rule diagrams, filaments, and percentile heatmaps. Their work also provides a description and design of the data mining and visual encodings to predict and explain longitudinal symptom development using the data of previous cancer patients including an evaluation by existing experts in the field.

They used a dataset of 699 patients with head or neck cancers where they recorded patient demographic information which contained quantitative, nominal, and ordinal data attributes (e.g., age, weight, disease stage, therapeutic combination). They also recorded longitudinal symptom data with quantitative values across time as a time-series attribute. The symptoms were grouped into 3 categories and then rated on a scale of 0-10 based on their severity.

On a high level, the tool simply aims to predict, filter, and explain longitudinal symptom development for cancer patients based on the data derived from existing previous patients using visual encodings like interactive association-rule graphs, filaments, and percentile heatmaps so that their system can be used to further research as well as in clinical therapy to provide patients with better course of treatment. They aim to make use of the "unreadable" lines data, cluster it using symptom cohorts and then find the best course of action for the individual patients based on their unique symptoms. The tool also allows them to run multiple scenarios on the data and test their hypothesis in real time while giving multiple customized visual encodings based on whatever parameters the researchers/physicians want to utilize.
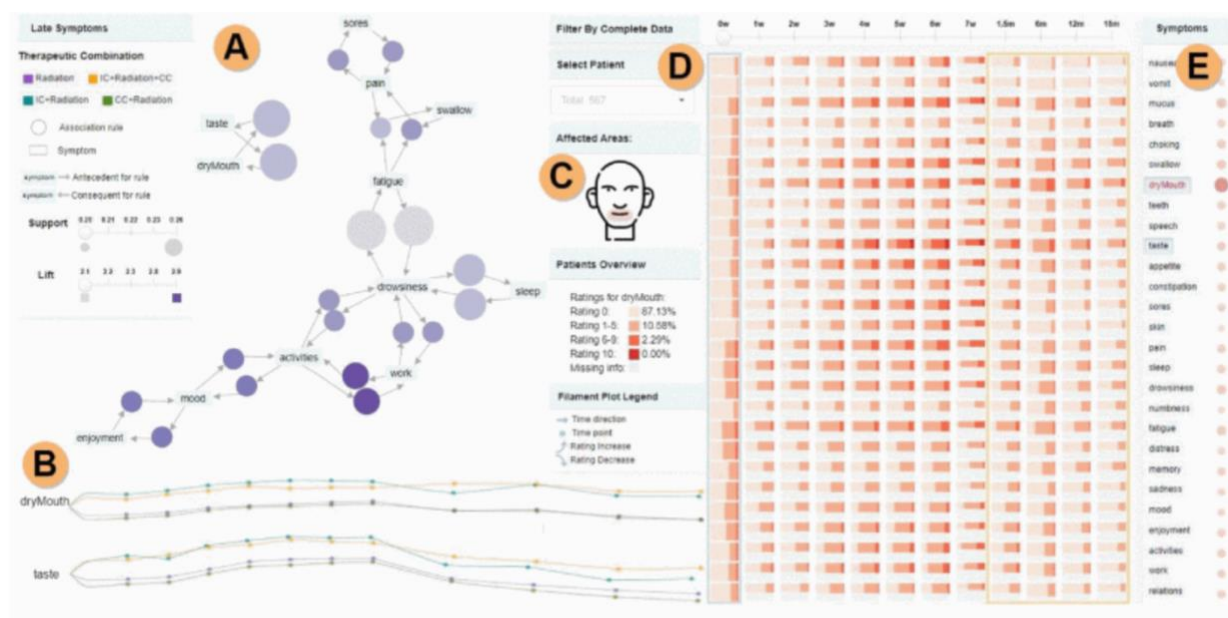
This is an image of one of the visual encodings provided by the tool. It uses association-rule-mining (ARM) relationships among the most frequent late-stage symptoms. Each bubble shows the rule and the size and color show us the support and life metrics. They use filament plots to show the symptom trajectory over time, visually color the affected areas, a correlation matrix, and a way to filter the data.

Filters

Therapeutic Combination
- ■ Radiation
- ■ IC+Radiation+CC
- ■ IC+Radiation
- ■ CC+Radiation

Rating Severity ×

Mild   Severe

Gender

● Female   ▼ Male

Tumor Category

• T0   ▪ T1   ▪ T2   ■ T3   ■ T4

Cluster by Symptoms ×

nausea   vomit   mucus   breath

choking   swallow   dryMouth

teeth   speech   taste   appetite

constipation   sores   skin   pain

sleep   drowsiness   numbness

fatigue   distress   memory

sadness   mood   enjoyment

activities   work   relations   walking

Scatterplots are used to show the parameters of the patient at a specific time during their treatment. These can be filtered based on the severity of the symptom and its designated cluster.

They use these encodings to analyze different phases in the therapy, how much the patients are affected (scale of 1-10) due to the symptoms, and symptom cluster diversity with how these symptom clusters are networked together and how they affect the patients.

They use a variety of marks and channels in their visual encodings. For the ARM relationships they use bubbles as points, as well as directed lines to show the correlation with other symptoms in the network. As for the channels, they utilize size and color of these points to give more information on the additional metrics, color saturation on the heatmaps to show the severity of the symptoms, spatial region and different sizes, shapes, and colors on the scatterplot to show the tumor category, gender and treatment provided.

	0w	1w	2w	3w	4w	5w	6w	7w	1.5m	6m	12m	18m	Symptoms

Symptoms
- nausea
- vomit
- mucus

Patients Overview

Ratings for mucus:
- Rating 0: 5.15%
- Rating 1-5: 56.64%
- Rating 6-9: 34.42%
- Rating 10: 3.79%
- Missing info:

mucus