

#### David Gotz, Jonathan Zhang, Wenyuan Wang, Joshua Shrestha, David Borland University of North Carolina at Chapel Hill

IEEE Transactions on Visualization and Computer Graphics, 2019

CPSC 547 | Kevin Chow

**Event Sequence Data via Dynamic** 

0.9



# **Event Sequences**

- Time-ordered lists of discrete events
- Analyze to discover patterns or rare event paths
- But... real-world datasets are large and complex:
  - Volume and length of event sequences
  - High-dimensional event data







### Volume and length of event sequences







### Volume and length of event sequences Aggregate sequences









### Volume and length of event sequences Aggregate sequences

### High-dimensional event data







![](_page_5_Figure_1.jpeg)

### Volume and length of event sequences Aggregate sequences

![](_page_5_Figure_3.jpeg)

Group events

![](_page_5_Picture_5.jpeg)

![](_page_5_Picture_12.jpeg)

# Grouping Events

- Typically, events are grouped in a pre-processing step
- Requires foreknowledge and expertise about events

![](_page_6_Figure_3.jpeg)

**I50: Heart Failure** 

. . . . . .

- Event type hierarchy
- ICD-10 Coding System

  - **150.2: Systolic Heart Failure** 
    - **I50.21: Acute Systolic Heart Failure**

![](_page_6_Picture_10.jpeg)

# Grouping Events

- Can't change event groups interactively
  - May want multiple groupings different levels of detail
  - An ideal grouping may not exist data- and taskdependent

![](_page_7_Picture_4.jpeg)

Attribu Size	rith outcome)				Non	nstraints e		
Race WHITE OR O BLACK OR A UNKNOWN OTHER RAC AMERICAN I ASIAN PATIENT RE NATIVE HAV	AUCASIA FRICAN / 2% E 1% NDIAN OI 1% 1% FUSED 0% WAIIAN OI 0%	34%	62%		Start of 365 da	letine		
Gender	9				Y			-
F		5	5%		=			
m		43%			ne			
Age					5			
	and the second se	in the second			ind			
0			103		¥			
Events								
Code	Description Evaluation and	Occurrence	s Patien	ts Correlation				
SNOMED- 30346009	established outpatient in office or other outpatient facility	17,101	1,563	-0.06				
SNOMED- 86181006	Evaluation and management of inpatient	14,909	1,576	0.00				
SNOMED- 12843005	Subsequent hospital visit by physician	12,583	1,419	.0.02				
SNOMED- 16310003	Ultrasonography	9,250	1,332	-0.09				
SNOMED- 118664000	Procedure on body system	8,871	1,548	-0.08				
SNOMED- 243120004	Regimes and therapies	7,723	1,231	-0.06				
SNOMED- 371571005	Imaging by body site	6.713	1,170	0.03		4		

![](_page_8_Figure_2.jpeg)

![](_page_8_Picture_3.jpeg)

![](_page_9_Figure_2.jpeg)

![](_page_9_Picture_3.jpeg)

Attributes Size 1,732 (7% with outcome)				Constrai None	ints		
Race   WHITE OR CAUCASIA   BLACK OR AFRICAN /   UNKNOWN 2%   OTHER RACE 1%   AMERICAN INDIAN OF 1%   ASIAN 1%   PATIENT REFUSED 0%   NATIVE HAWAIIAN OF 0%	34%	62%		Start of 365 da	2		
Gender	5	5%		ay ti			
M	45%			Be			
Age				<u>₹</u> .			
	all this	-		<u>d</u>			
0		103		¥			
Events							
Code Description Evaluation and management of	Occurrence	s Patien	ts Correlation				
SNOMED- 30346009 outpatient in office or other outpatient facility	17,101	1,563	-0.06				
SNOMED- 86181006 Evaluation and management of inpatient	14,909	1,576	0.00				
SNOMED- Subsequent hospita 12843005 visit by physician	12,583	1,419	-0.02				
SNOMED- 16310003 Ultrasonography	9,250	1,332	-0.09				
SNOMED- Procedure on body 118664000 system	8,871	1,548	-0.08				
SNOMED- Regimes and 243120004 therapies	7.723	1,231	-0.06				
SNOMED- 371571005 Imaging by body site	6.713	1,170	0.03				

![](_page_10_Figure_2.jpeg)

![](_page_10_Picture_3.jpeg)

Attribu Size 1,732 (7% w Race white or o BLACK OR A UNKNOWN OTHER RAC AMERICAN I ASIAN PATIENT RE NATIVE HAV Gender F M Age	rith outcome) CAUCASIA FRICAN / Z% E 1% NDIAN OI 1% 1% FUSED 0% NAIIAN OI 0%	34%	62% 5%	
Events	Description	Occurrence	s Patien	ts Correlation
SNOMED- 30346009 SNOMED- 86181006	Evaluation and management of established outpatient in office or other outpatient facility Evaluation and management of	17,101	1,563	-0.06
SNOMED- 12843005 SNOMED-	Inpatient Subsequent hospital visit by physician	12,583	1,419	
16310003 SNOMED- 118664000 SNOMED- 243120004	Procedure on body system Regimes and therapies	9,250 8,871 7,723	1,332	-0.08
SNOMED- 371571005	Imaging by body site	6.713	1,170	0.03

![](_page_11_Figure_2.jpeg)

![](_page_11_Picture_3.jpeg)

![](_page_12_Picture_1.jpeg)

events based on an informativeness score

# Determining an optimal and adjustable level of grouping

![](_page_13_Picture_3.jpeg)

- I. Determining an **optimal** and **adjustable** level of grouping events based on an **informativeness score**
- Supporting navigation of the event type hierarchy with a scatter-plus-focus visualization

11

- I. Determining an **optimal** and **adjustable** level of grouping events based on an **informativeness score**
- Supporting navigation of the event type hierarchy with a scatter-plus-focus visualization
- 3. Scenting to enable discovery of interesting event types

![](_page_15_Picture_4.jpeg)

## Informativeness Score

- Computed for each event type *j* in the event type hierarchy
- Measures the strength of the association between an event type and the outcome
  - If this patient had outcome v, did they also experience event type j?
- Based on the chi-square te

![](_page_16_Picture_5.jpeg)

est statistic 
$$X_j^2$$

![](_page_16_Picture_7.jpeg)

## Algorithm: Optimal Grouping Level

- <u>Goal</u>: Determine the most informative cut through the event type hierarchy
- Recursively traverse event type hierarchy
- Compare informativeness score of parent with each child

![](_page_17_Picture_4.jpeg)

![](_page_17_Picture_5.jpeg)

![](_page_18_Figure_1.jpeg)

K controls level of aggregation (larger = more aggregation)

![](_page_18_Picture_7.jpeg)

## Scatter-plus-Focus

![](_page_19_Figure_1.jpeg)

Scatter plot

![](_page_19_Figure_3.jpeg)

#### Focused dual-view

![](_page_19_Picture_5.jpeg)

## Scatter-plus-Focus

- Challenges of overplotting!
- Grey hexes hint at density of all possible event types
- Marks are only event types part of informative cut
  - Control R with slider

![](_page_20_Figure_5.jpeg)

![](_page_20_Picture_6.jpeg)

## Scatter-plus-Focus

- Focuses on hierarchy of selected event type
- X-axis is centred on correlation
- Y-axis: determined by optimization-based layout algorithm

![](_page_21_Picture_4.jpeg)

#### **Event-Outcome Associations**

![](_page_21_Figure_6.jpeg)

![](_page_21_Picture_8.jpeg)

# Algorithm: Optimize Layout

- Cost function that balances two layout priorities:
  - Y-positions should be close to original in scatter view Marks should not overlap
- Two constraints:
  - Optimized y-positions must be within y-axis scale Original y-position order of marks must be preserved

![](_page_22_Picture_7.jpeg)

# Algorithm: Optimize Layout

#### No changes to y-positions

![](_page_23_Figure_2.jpeg)

#### With algorithm

![](_page_23_Figure_4.jpeg)

![](_page_23_Picture_5.jpeg)

## Scenting

- Size of glyph indicates magnitude of scent value

![](_page_24_Figure_4.jpeg)

## Shows up when exploring type hierarchy in focused view Scent value: range of correlations to outcome in children

![](_page_24_Picture_6.jpeg)

### Evaluation

- experience
- Results from thematic analysis:
  - Training is required
  - Automated selection of aggregation level useful
  - Navigating through event type hierarchy was intuitive

#### • 3 medical experts: health researchers with data analysis

# Hands-on demonstration and semi-structured interviews

![](_page_25_Picture_9.jpeg)

## What-Why-How Analysis

#### What: Data

- Tree (event type hierarchy)
- Table (patient data)
- What: Derived
- Optimal event grouping
- Informativeness score, scent value, optimized ypositions

![](_page_26_Picture_7.jpeg)

#### <u>Why</u>

### Discover and produce (event type groupings)

Scale: 5,000 patients, 700,000 events, 10,000 unique event types

![](_page_26_Figure_11.jpeg)

![](_page_26_Figure_12.jpeg)

![](_page_26_Picture_13.jpeg)

![](_page_26_Picture_14.jpeg)

## What-Why-How Analysis

#### How: Encode

- Scatterplots
- Color (outcome correlation)

#### How: Reduce

- Item aggregation (grouping event types)
- Scenting (picking event type)

![](_page_27_Picture_7.jpeg)

### How: Change

- Select (mark in scatter)
- How: Facet
  - Overview+detail view (scatter-plus-focus)
  - Layering (grey hexes in background)

![](_page_27_Picture_13.jpeg)

## Critique

- Strengths
  - Intuitive, simple algorithms

  - load
  - Generalizable to contexts other than health

#### Dealt with challenges of occlusion and distortion

#### Switching between views and parameter control reduces

![](_page_28_Picture_10.jpeg)

# Critique

- Weaknesses/Limitations
  - custom groupings
  - Adding event type groups can be tedious
  - Reliance on tree-based event type hierarchy

Automated approach to aggregation may hide better

![](_page_29_Picture_7.jpeg)

![](_page_30_Figure_1.jpeg)

#### Visual Analysis of High-Dimensional Event Sequence Data via Dynamic Hierarchical Aggregation

![](_page_30_Picture_3.jpeg)

![](_page_30_Picture_4.jpeg)