Unsupervised Wisdom: Explore Medical Narratives on Older Adult Falls

Mbamali Ifechukwu

1. Abstract

The analysis made use of *embeddings*, *dimensionality reduction*, *clustering algorithms*, *network graphs and text summarization* techniques to effectively identify, and understand themes from medical narratives on older adults falls.

Key findings: DBSCAN outperforms k-means in cluster identification. Patients in the "*Alcohol-Related Head Injuries and Falls*" group tend to be younger, while the "*Atrial fibrillation related falls*" group was generally older, and the "*Syncope-Related Head Injuries*" group had a higher rate of severe cases compared to others. In comparison to the previous year, cases involving "*Head Injuries from Falls*", "*Syncope-Related Head Injuries*" and "*Rib Injuries from Falls*" saw the most significant increase in the average number of cases.

2. Methodology

a. Datasets and Preprocessing: The analysis utilized two datasets, the primary dataset and the OpenAI embeddings dataset. Some of the general preprocessing steps include:

(i) New fields: Additional columns were introduced to the primary dataset[1] to determine severity levels based on disposition column and another to categorize activities based on the narratives.

(ii) Replacing Abbreviations: The narrative column was also processed by replacing abbreviations with full clinical definition to improve readability.

(iii) High dimensionality reduction: The embeddings data (115128 by 1537), underwent preprocessing including normalization and dimensionality reduction via principal component analysis (PCA) to achieve a reduced dimension (115128 by 151). The aim was to reduce computational requirements while preserving crucial information.

b. Analytics Approach: The methods include:
(i) Clustering: The analysis employed k-means and DBSCAN algorithms in conjunction with UMAP processing on the dimension-reduced (PCA) embeddings data to uncover key themes cluster based on narrative embeddings.

(ii) Network graphs: these were used to explore keyword pair occurrences within narratives for different clusters, with keyword ranking via the PageRank algorithm highlighting significant terms. processing steps used here include body-part and stop-words removal.

(iii) Text ranking and summarization: These methods were applied to generate summaries for narrative clusters to provide insights into key themes. **Limitation:** Due to size of data in each cluster, random sampling technique was used to reduce the size of data to be summarized for computational reasons.

3. Results and Insights

(i) Effectiveness of Embeddings: Using embeddings proved effective in extracting themes.

(ii) DBSCAN vs. K-means: Density-based clustering (DBSCAN) demonstrated greater efficacy in identifying clusters compared to the k-means approach.

(iii) Theme Identification: Density-based clustering revealed the presence of 9 major themes, with cluster 0 being categorized as outlier/general themes.

(iv) Text Summarization: The text summarization algorithm effectively condensed core narratives within clusters, enabling the description of associated themes for each cluster.

(v) Theme Understanding: see table in the section below. Patients in cluster 6 are generally younger, while those in cluster 8 are older compared to the rest of the group. Additionally, cluster 4 has a higher prevalence of severe cases compared to other groups.

Unsupervised Wisdom: Explore Medical Narratives on Older Adult Falls
Mbamali Ifechukwu

4. Tables and Visualizations
a. Tables

Cluster	Theme	Associated Activities	Obstacle	Injury	Top 3 Keywords (excluding the term "Fall")
0	General Elderly Falls and Injuries	Lost Balance	Ladders, others not specified	Others	left, admit, contusions
1	Head Injuries from Falls	Standing, Rising	Bed or bedframes	Laceration	injury, laceration, contusion
2	Falls Resulting in Shoulder Injuries	Tripped, Exercise, Sports	Exercise	Dislocation, Avulsion, Strain & Sprain	fracture, left, humerus
3	Hip Injuries from Falls	Rising, Tripped	Footwear	Fracture, Strain & Sprain	fracture, left, femur
4	Syncope-Related Head Injuries	Fainted	Toilets	Laceration	syncope, laceration, striking
5	Rib Injuries from Falls	Standing	Bathtubs or Showers	Fracture	left, fracture, ribs
6	Alcohol-Related Head Injuries and Falls	Stair Navigation, others	Stairs or steps	Poisoning, Laceration	alcohol, blood, intoxication
7	Buttocks Contusions from Falls	Rising, sitting, slipped	bed or bedframes	Contusions	contusions, buttocks, lower
8	Atrial fibrillation related falls	Sitting, Standing	Tables, rugs & carpets, Ceilings & Walls	Hematoma	encounter, laceration, initial
9	Floor Falls and Associated Injuries	Walking, Slipped	Floors, balconies	Contusions	Falling, Floor, Dizzy

Table 1: The table above captures the themes associated with each cluster.

Unsupervised Wisdom: Explore Medical Narratives on Older Adult Falls
Mbamali Ifechukwu

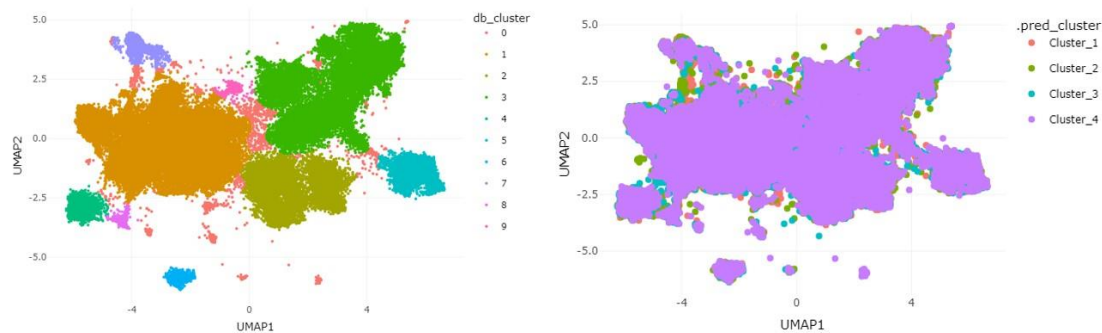## b. Visualizations

DBSCAN vs K-Means

Figure 1: L-R The result of density-based clustering vs k-means

Understanding Themes

Figure 2: L-R Activities leading to falls based on theme-cluster 6; keywords associated with theme-cluster 6

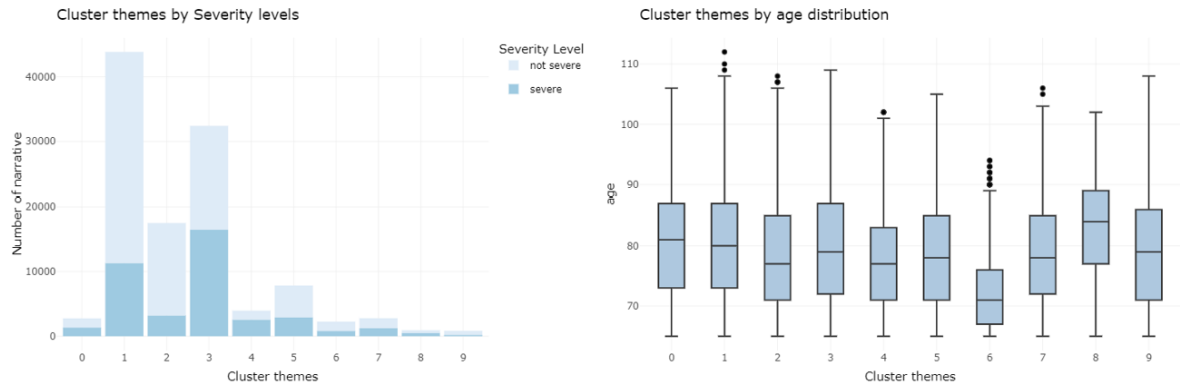### Exploring themes with relation to other variables

Figure 3: L-R Cluster themes by number of narratives & severity levels; cluster themes by age