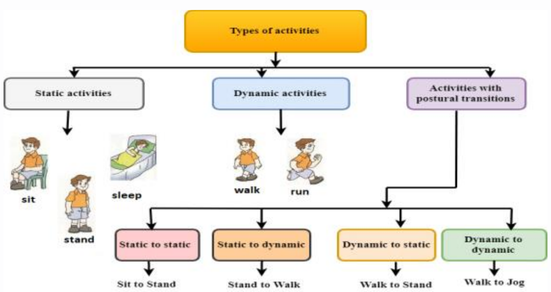
Unsupervised Wisdom: Midpoint Summary

This is a midpoint summary of our ongoing work on the open unsupervised wisdom competition. So far, we have restricted to exploring the primary and secondary CSV files provided and have not evaluated external data. While much of the narrative of a fall incident has been extracted and stored in the columns, the “precipitating event” is not. While not all narratives give this informative description, there are nevertheless many fall incident reports that do show this information explicitly or implicitly. Our work has been concentrated on extracting and analyzing this ‘precipitating event’, from simple methods to LLM usage. We also explored appropriate nodes and edges that are useful in creating a knowledge graph and informative visualization.

Figure 1



As such we describe our design for extracting the precipitating event. Activities, in general, can be categorized as shown in Figure 1, so a simple ‘keyword search’ for ['sitting', 'sleeping', 'walking', 'getting up', 'getting down', 'standing', 'climbing', 'running', 'dancing'] activities gives us 16000 samples. A histogram for a subset of activity (Figure 2) shows this. Similarly, supplementary data shows a similar distribution. We further do a ‘semantic search’ with FAISS(Facebook Similarity Search), using OpenAI’s text-embedding-ada-002 model provided by the organizers and the activity category as a ‘range search’ query. In Figure 3 we clearly see more samples for ‘sleeping’, which can be explained by the fact that semantic search matches narratives such as falling from bed’ to ‘sleeping’ where a simple keyword search fails.

Figure 2

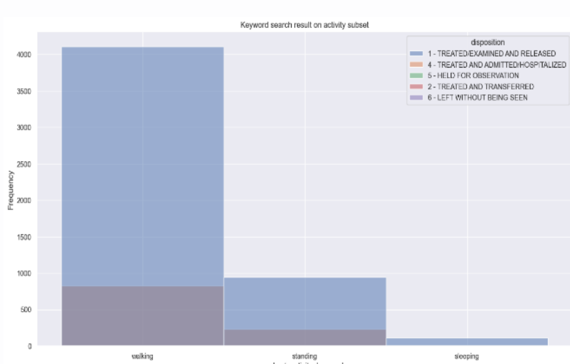
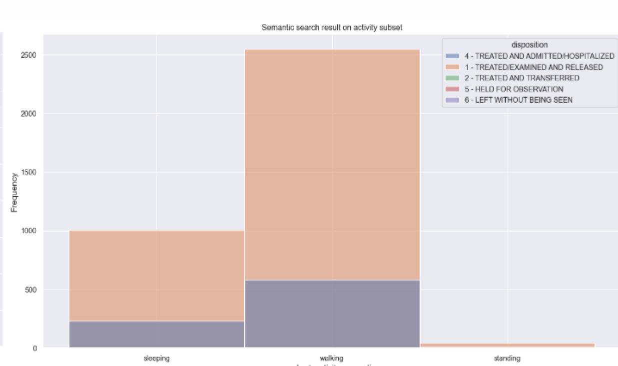


Figure 3



We also explored a prompt-based text-generation approach using calls to OpenAI's ChatGPT3.5 using the narratives. The exact and striped response extracted are far richer in the context that the returned activity described are more fine-grained such as 'sweeping and falling' or 'rolling off the bed' but the response time for OpenAI calls is time-consuming and limited by rate limits, so we only extracted ~6000 samples. We then use the refined response thus obtained to compute embeddings using Falcon-7B locally and then clustered the result using TSNE as shown in Figure 4. This is promising but further exploration using this setup will require some thought due to OpenAI's response time as mentioned. We evaluated the prompt-based LLM text-generation approach using Falcon_7B locally (quantizing to 4bit) but this is also slow but could be improved to an acceptable time frame with higher-end GPU.

Figure 4

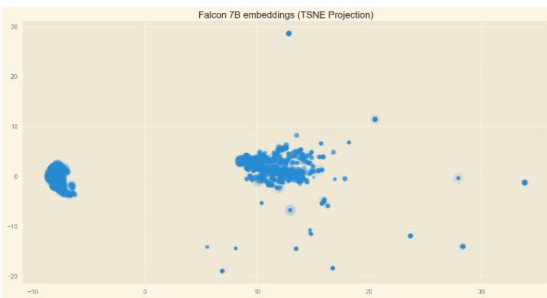
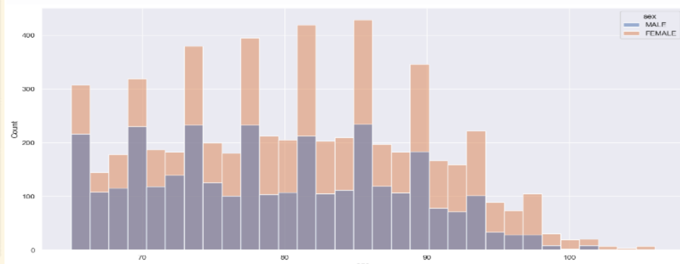
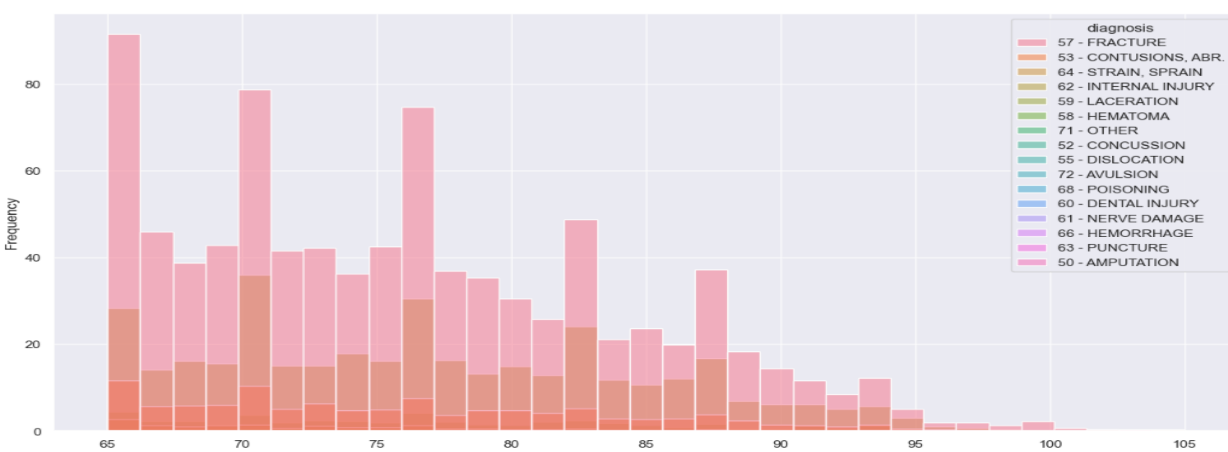


Figure 5

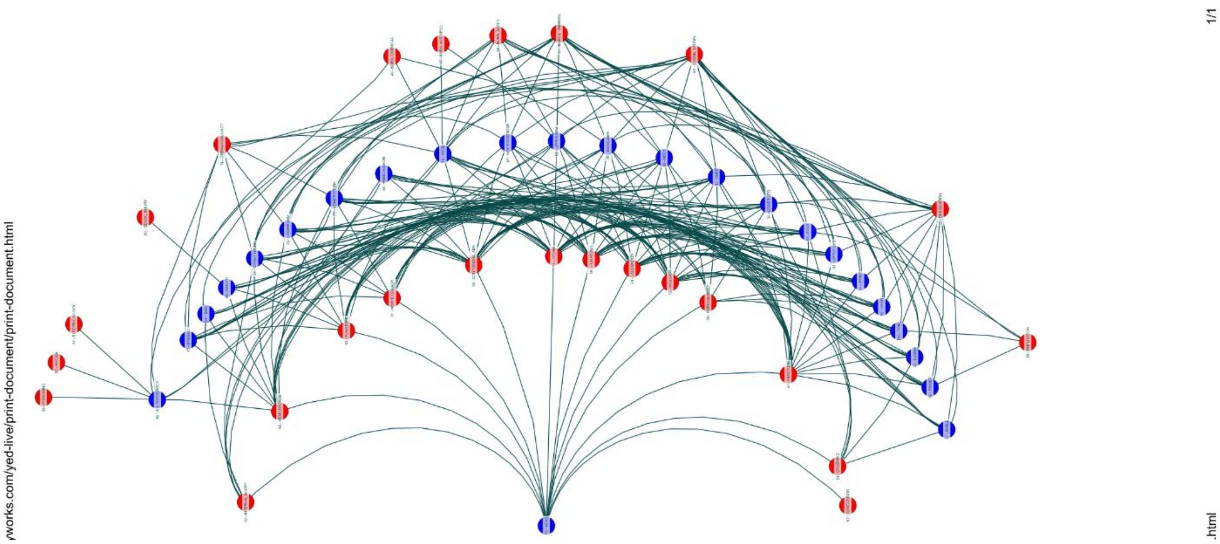


The location columns do not categorize the places so that the event can be separated to occur in toilet/bathroom/restroom or not. This information though can be obtained from the product columns. Figure 5 shows female falls in the toilet increases in comparison to male as age progresses.

Falling on stairs, roofs, or from some height can be more hazardous and lead to a higher risk of fall fatality. We extract these incidents from the narrative or product column and explore their impact on fall severity (diagnosis). The figure below shows fracture to be more common than abrasion or strain/sprain in such cases.



Going forward, we plan on increasing the accuracy, as well as the relevant number of samples, of unsupervised precipitating event class from the narrative. Some narratives imply the event, and the activity category needs to be reasoned out, perhaps with a 'chain of thought' based prompt approach. Activities with postural change as in Figure 1 could also lead to meaningful insight. We preferably want to do all these with Falcon-7B locally. We have explored knowledge graphs interactively as shown below connecting 'body parts' to diagnosis; we are continuing to explore the edges that can best describe the precipitating event.



Some Key Findings and Future Direction

- Keywords search is not enough to extract precipitating events from the narrative if given but can be a useful benchmark in comparing and validating other approaches. Semantic search can return relevant result class but can also return more false positives specifically when no precipitating event is explicitly mentioned.
- OpenAI's ChatGPT3.5 returns refined and pertinent activity or not mentioned if not mentioned but is slow and rate limited. Falcon-7b though can be run locally, is still slow on lower-end GPU and the results are not as clean as OpenAI's but this could be worked on.
- Falling from heights such as stairs, and roofs are more hazardous leading to higher severity of falls. We feel more semantic search or LLM-based approach on the narratives is necessary here.
- Fall in Toilet/restroom or shower, slipping in water as opposed to just slipping has its own characteristic. This needs further re-evaluation.
- Graph with appropriate template shows better data visualization and relations. This is though better suited to interactive notebooks.