Process Mining Case Study

We present a case study for the use of Process Mining techniques on an event log assembled from Category 2 Job Cycle data ranging between August and December 2022 from the East Midlands ambulance service. The original data had data quality problems with misreported and unnecessary events (Figure 1). Applying the business rules to the data and using Directly Follows Graphs enabled the observation of the two most common recorded traits followed by ambulances responding to category 2 data, those who were seen and treated and those who were treated and conveyed (Figure 2).A white and blue diagram

Description automatically generated with medium confidence

Figure 1: A directly follows graph of the unfiltered event log including 152,525 patients’ Job Cycles with

A diagram of a flowchart

Description automatically generated

Figure 2: A Directly Follows Graph (DFG) showing the top 2 traces accounting for 99% of the Job Cycle  
traces. The 2 traces show patients who were treated and transported to the hospital and those patients who were only treated at the scene. The remaining 1% of traces not shown here were missing some of the activities shown in this DFG due to issues in their recording.

Conformance techniques were used to measure how well the mined process models fitted the event log data. However, the variation in the treatment activities also caused the mined models to allow events including the call for an ambulance to happen in a sequence counterintuitive to real ambulance Job Cycles (Figures 3 and 4).

A black and white image of small squares

Description automatically generated

Figure 3: A petri-net process model produced using the Heuristic mining algorithm on the event log. The event log had patients with anomalously ordered events removed but the events recording treatments and vital signs measurements were left in so that the miners could show concurrency. This petri-net does allow both the First and Last NEWS to be measured prior to the Incident being called which is behaviour not seen in the event log but it does ensure that the key events documenting the progression of the ambulance through the Job Cycle happen in the expected order.

A diagram of a diagram

Description automatically generated with medium confidence

Figure 4: A petri-net process model produced using the Inductive mining algorithm on the event log. The event log had patients with anomalously ordered events removed but the events recording treatments and vital signs measurements were left in so that the miners could show concurrency. This petri-net allows behaviour that cannot happen in the real world like the incident the ambulance is responding to happening, at a later time than the patient being transferred to the hospital. Allowing a lot of behaviour not seen in the event log makes this process model imprecise.

We were able to obtain more date during the project extending our data from August 2022 to August 2023. Conformance techniques were also able to show a change in process (known as concept drift) during the winter months of 2022 compared to the process model mined over the whole August 2022 to August 2023 period (Figure 5).

A graph with numbers and lines

Description automatically generated

Figure 5: A graph showing concept drift or the change in a process over time. A process model was created from an event log of Job Cycle data for the entire year between August 2022 and August 2023 using the heuristic miner. This graph shows the fitness of each months data to the process model created from the entire year’s data. The drops in fitness during the winter months of 2022 indicate an increase in behaviours not playable in the model.

In our Enhancements investigation the timings between key elements of the ambulance Job Cycle were measured with respect to different attributes from the event log to infer their effect on the process (Figure 6).

A graph of a graph

Description automatically generated with medium confidence

Figure 6: Boxplots showing the distribution in the time spent on the scene for category 2 calls with respect to the highest qualification of the health care team on scene. ECA and ECP stand for emergency care assistant and emergency care practitioner respectively while UCA stands for urgent care assistant. The anomaly of the UCAs’ increased time spent on scene was explained by EMAS stating that the UCAs often wait for additional support from other health care professionals.

A DFG showing the timings between events was produced to observe bottlenecks (Figure 7).

A diagram of a flowchart

Description automatically generated

Figure 7: Directly Follows Graph showing the top 2 traces accounting for 99% of the Job Cycle traces.  
The 2 traces show patients who were treated and transported to the hospital and those patients that were only treated on the scene. The remaining 1% of traces not shown here were missing some of the activities shown in this DFG due to issues in their recording. A bottleneck can be observed between the call being received (Incident) and an ambulance crew being allocated (Originated) and another bottleneck between the ambulance arriving at the hospital (Arrive Dest) and patient handover (Care Transfer)

Machine learning models were trained to predict if a patient was to be transported to the hospital with the secondary aim to understand the influence certain attributes had on the prediction (Table 8).

Table : The sum of the importances for the different categories of input features used in a Decision Tree to  
predict whether a patient was to be taken to hospital using data available from early on in an ambulances  
Job Cycle.

|  |  |
| --- | --- |
| **Decision Trees Features** | **Feature Importance** |
| Initial Clinical Category of Call | 0.29 |
| First Recorded National Early Warning Score (NEWS) | 0.23 |
| Age of Patient | 0.17 |
| Hour of Arrival on Scene | 0.10 |
| Index of Multiple Deprivation (IMD) Decile at Incident Location | 0.075 |
| Day of the Week | 0.065 |
| Highest Qualification on Scene | 0.038 |
| Gender of Patient | 0.022 |
| Location of Incident (Home/Other) | 0.015 |
| Whether specific Job Cycle activities have occurred prior to arrival on the scene | 0.0013 |

In conclusion, we followed the PM2 process methodology and used Discovery, Conformance, and Enhancements techniques on Category 2 Job Cycle data provided by East Midlands Ambulance Service. Data quality issues were tackled with business rules defined with help from clinical experts at EMAS. The 2 most common Job Cycle traces were observed in the data with bottlenecks in the processes identified. The effect on attributes recorded in the event log on the time between specific events were measured. Future work should focus on measuring how pathways, activities and attributes effect a patient's survival to ensure we are providing the best care to patients