Application of Bayesian Networks

Matthew Fleischman (FLSMAT002)
Ariel Levy (LVYARI002)

October 2024

1 Introduction

1.1 Background

Traffic accidents pose significant challenges globally, impacting public safety, straining healthcare systems, and hindering economic growth [1]. Analyzing data from specific regions, such as the UK between 2012 and 2014, can provide valuable insights into the causes and patterns of road incidents. When large datasets are available, these findings can be generalised to inform traffic safety measures in other regions, especially those with an infrastructure quality comparable to that of the UK.

1.2 Objective

With this in mind, the objective of this analysis is to develop a tool that can be used to determine the necessary level of emergency response that will bring about a reduction in traffic accident severity and casualties. The ultimate goal is to improve road safety by decreasing the incidence of fatalities or serious injuries in accidents that do occur.

1.3 Potential User Community

This tool and its insights will be useful to a variety of stakeholders, including:

- **Governmental Policy Makers:** To make informed decisions on road safety regulations.
- **Transportation Authorities:** To identify accident hotspots and improve road designs.
- **Public Transport Providers:** To plan safer, less accident-prone routes.
- **Traffic Police:** To optimise the deployment of patrols and mitigate the severity of incidents.
- **Health Officials:** To assess the impact of road accidents on the healthcare system and deploy ambulances promptly when required.
- **Automotive Manufacturers:** To enhance vehicle safety measures.
- **Insurance Companies:** To produce fair and accurate risk assessments and thus pricing.

2 Problem Analysis

2.1 Dataset and Factors

A dataset containing all traffic accident data between the years 2012 and 2014 was used to construct a Bayesian – and thereafter – decision network [2]. Out of the 128 columns in the dataset, nine were selected as critical for supporting the objectives of this project. The important factors identified include:

- **Urban or Rural:** Differentiates between urban and rural areas, as traffic patterns and accident causes may vary between these environments.
- **Road Type:** Identifies the type of road, which influences traffic behaviour.
- **Number of Vehicles Involved:** Indicates how many vehicles were part of the accident.

- **Number of Casualties:** Represents the total number of injuries and fatalities resulting from the accident.
- **Speed Limit:** Details the maximum legal speed for the road where the accident occurred.
- **Road Surface Conditions:** Describes the condition of the road surface, which can influence vehicle handling.
- **Light Conditions:** Refers to the visibility at the time of the accident, affecting driver perception and reaction times.
- **Weather:** Encompasses various meteorological conditions that can impair visibility and affect road traction.
- **Accident Severity:** Classifies the seriousness of the accident.

2.2 Expert Knowledge

This dataset is large ($\sim$ 450000 entries) and comprehensive, minimising the need for extensive expert knowledge, particularly concerning probabilities. Nonetheless, papers like [3] offer valuable insights into factors that lead to higher casualty accidents and, thus, will be used to inform the utilities used in the decision network.

3 Decision Network Model

A diagram of the model can be seen in Figure 1. The decision network has nine chance nodes that consider the variables described in Section 2.1. The *area of the accident*, *weather conditions*, and *light conditions* are independent. The conditional probability distributions of the other chance nodes are calculated by cross-tabulating and normalizing the data accordingly. The *number of vehicles*, *speed limit*, *road surface conditions*, and *light conditions* are considered by the *level of emergency response* decision node. The utility node considers the *number of casualties* (C), *accident severity* (S), and *level of emergency response* (R), which each have a domain of $\{1, 2, 3\}$. The utility table is populated by the values calculated with Equation 1. These utility values are computed for each permutation of the parent nodes' possible values.

$$U = 3 - |R - \frac{C+S}{2}|. \tag{1}$$

This utility equation rewards response levels that are more closely aligned with the accident severity and number of casualties. For example, R=3, C=3, S=3 provides a utility value of 3. Whereas, R=1, C=3, S=3 only provides a utility value of 1.

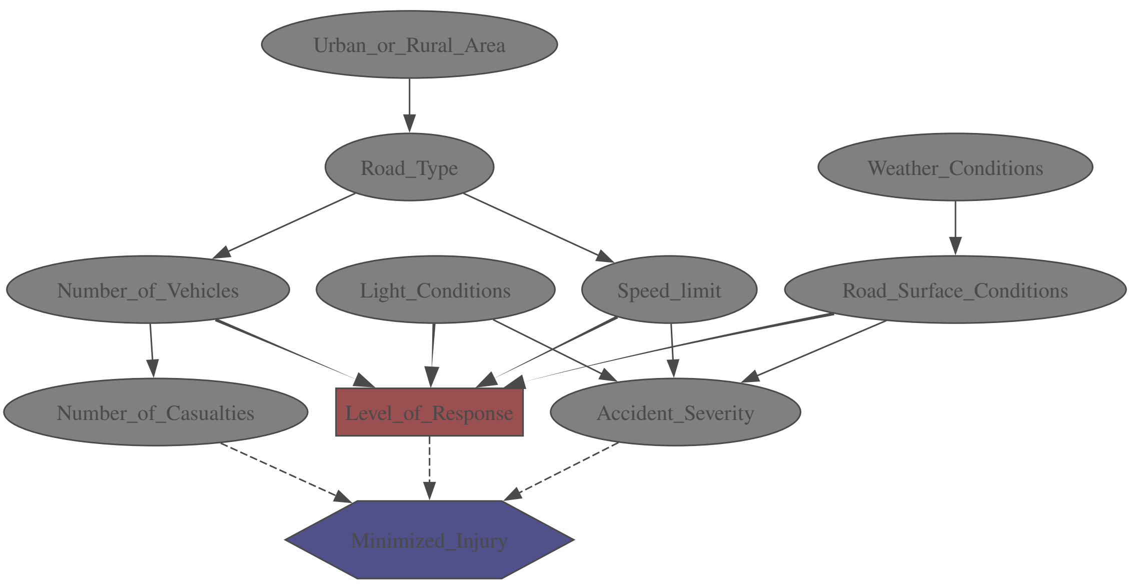

Figure 1: Diagram of Decision Network Structure.

4 Model Testing and Evaluation

4.1 Inferences on Unobserved Variables

Three steps were used to perform inferences on unobserved variables. First, the inference engine was set to the LIMID inference method:

```
ie=gum.ShaferShenoyLIMIDInference(bn)
```

Second, any available evidence was supplied and inferences were made:

```
ie.setEvidence({evidence_var: value})
ie.makeInference()
```

Finally, the probabilities of the desired unobserved variable were shown by calling:

```
gnb.showProba(ie.posterior(target_var))
```

As an example, the *number of vehicles* node was given the evidence value of 3, and the *number of casualties* node was inferred to produce the probability distribution shown in Figure 2

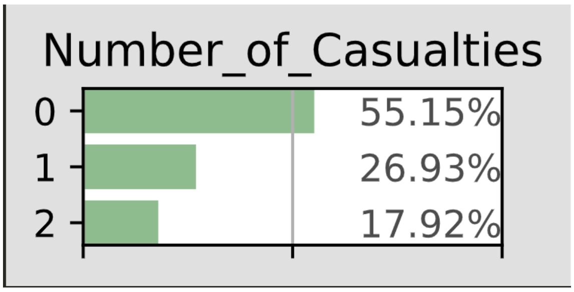

Figure 2: Inference of Number of Casualties Given 3 Vehicles Present.

4.2 Decision Network

Any available evidence was given to the inference engine as described in Section 4.1. The probabilities of all unobserved chance nodes were inferred, and the decision node was set to make the optimal decision by calling:

```
decisions = ie.optimalDecision("Level_of_Response")
```

This produced a table of the responses determined by the decision node corresponding to all permutations of the possible values for *light conditions, road surface conditions, speed limit,* and *number of vehicles*; one can either observe or infer the probabilities of the decision node's parents. Once the values for these nodes are selected, the appropriate level of response can be found in the decision table.

Various tests were performed to ensure the decision network was behaving as expected. Here is one example: The weather condition was set to *fine without high winds*, the area type was set to *urban*, and the resulting decision table was returned. A portion of this table is shown in Figure 3. If we consider the ideal lighting and road surface conditions, the slowest speed limit, and the fewest number of vehicles involved, the values in the variable columns should be (0, 0, 0, 0). Tracing these values to the *level of response* column, we get the recommended response of 0. This was expected.

This decision network ultimately aligns with the objective of minimising harm associated with traffic accidents, by analysing various conditions (light, road surface, speed limit, vehicle count, etc.) to determine the optimal emergency response. Based on these factors, traffic and health services can prepare for and respond to accidents with the appropriate urgency and resources, thereby minimising damage and reducing casualties.

				Level_of_Response		
Light_Conditions	Road_Surface_Conditions	Speed_limit	Number_of_Vehicles	0	1	2
		0	0	1.0000	0.0000	0.0000
			1	1.0000	0.0000	0.0000
			2	1.0000	0.0000	0.0000
			3	1.0000	0.0000	0.0000
			4	1.0000	0.0000	0.0000
			5	1.0000	0.0000	0.0000
			6	0.0000	0.0000	1.0000

Figure 3: Portion of Decision Table.

5 Conclusions

This research designed and implemented a decision network using PyAgrum to optimize emergency response levels in traffic accidents. First, a Bayesian network with 9 chance nodes was created. The values of the chance nodes could either be determined by evidence or inference. A decision node and utility node were then added to create the final decision network. The decision network allowed for appropriate emergency response levels to be recommended.

This model is adaptable and can be easily modified to work off of different datasets for different regions. Future work could expand the model, refine the utility calculations, or explore optimising the datasets supplied to the network for more accurate results.

References

[1] World Road Association (PIARC). Road safety manual: Strategic global perspective: Scope of the road safety problem, 2024. URL https://roadsafety.piarc.org/en/strategic-global-perspective/scope-road-safety-problem. Accessed: 28-09-2024.

[2] Dave Fisher-Hickey. 1.6 million uk traffic accidents, 2017. URL https://www.kaggle.com/datasets/daveianhickey/2000-16-traffic-flow-england-scotland-wales. Accessed: 28-09-2024.

[3] Rodger Charlton and Gary Smith. How to reduce the toll of road traffic accidents. *Journal of the Royal Society of Medicine*, 96(10):475–476, 2003. URL https://doi.org/10.1177/014107680309601001.