Artificial Intelligence
Università della Svizzera Italiana
Faculty of Informatics
January 9, 2020

Assignment 2: Causal Inference

Giorgia Adorni (giorgia.adorni@usi.ch)

1 Structure of the Network

The problem modelled is a generic trip by car, influenced by factors such as the strike of public transport, road works or even the weather, and the delay that comes with it. The causal diagram that model this problem includes the following variables:

- `Weather`: weather during the journey that should be *sunny* or *rainy*.
- `Strike`: *true* if a strike of public transport takes place, *false* otherwise.
- `RushHour`: *true* if the time it's rush hour, *false* otherwise.
- `RoadConditions`: condition of the road floor, that depends on the weather, and should be *dry* or *wet*.
- `Humor`: humor of the driver, dependent on the weather, that is *good* or *bad*.
- `RoadWorks`: *true* if there are road maintenance works in progress, *false* otherwise. This variable depends on the conditions of the road.
- `Speed`: driving velocity, that should be *slow* or *fast*, dependent on the road condition, if it's rush hour and if there are road works.
- `Danger`: danger incurred during the trip, that should be *low* or *high*, dependent on the driving velocity, the road conditions and if it's rush hour.
- `Accident`: accident risk, that could be *low* or *high*, influenced by the danger, the humour of the driver and if there is a strike of the public transport.
- `Delay`: *true* if the trip is delayed, *false* otherwise.

The objective of the network is highlight hoe weather and humour impacts on travel safety. The graph could provide valuable indications about the correlation between the driver humour and a delayed trip, or for example between the weather and the risk of an accident and also on how the road conditions influences car crashes.

Each node is connected by an arrow to one or more other nodes upon which it has a causal influence. Most of the arcs orientation are self-explaining. An exception was made for the `Humor` variable, that influences the accident risk and is caused only by the weather and not for example by the delay.

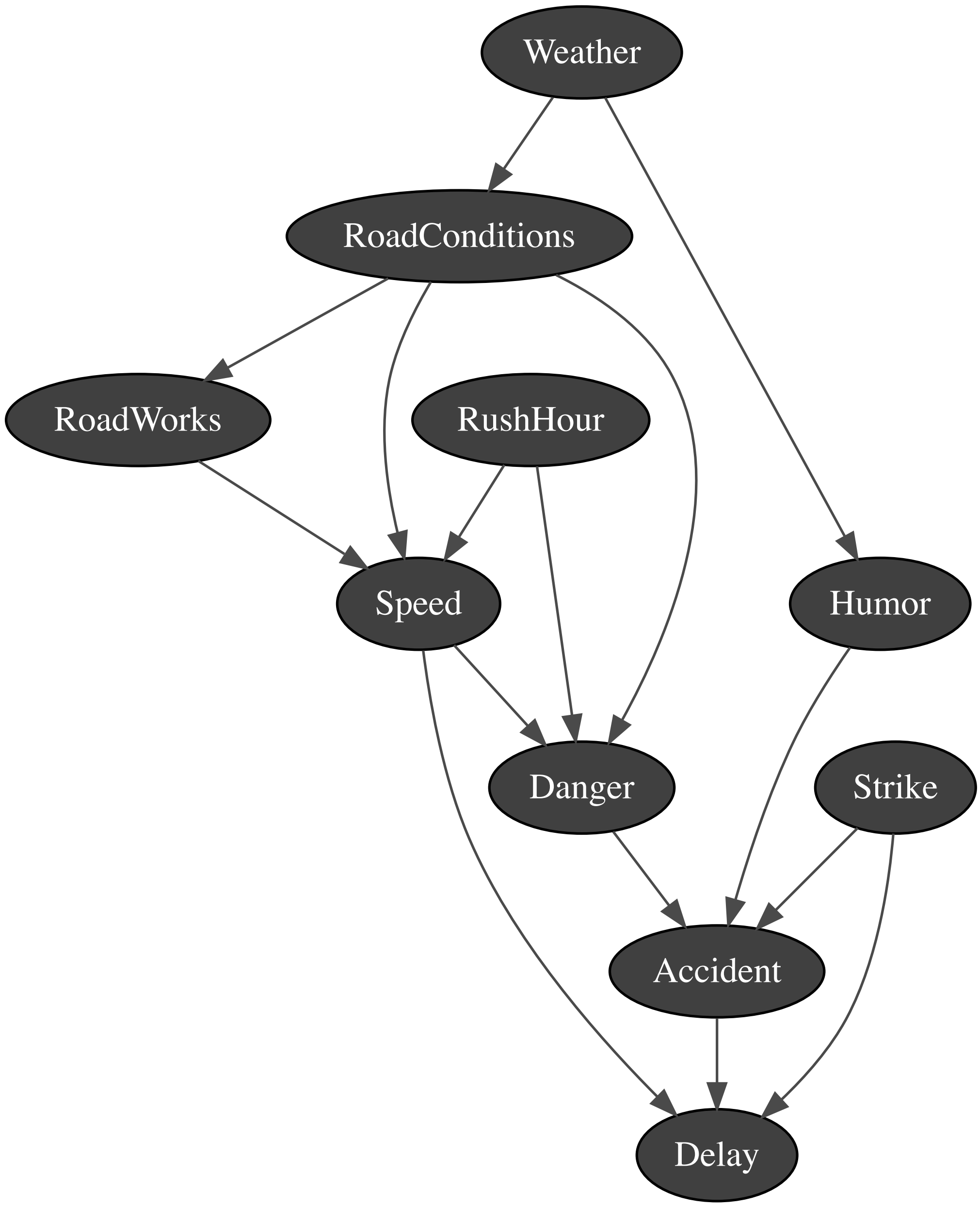

Figure 1: Bayesian Network

An arc can be inverted if and only if no v-structures, i.e. colliders in which the parents are not adjacent, are generated or destroyed in doing so.

In the model there are only two arrow that can be inverted, and these are the one that goes from the variable `Weather` to `Humor` and the one that goes from the variable `Weather` to `RoadConditions`. In fact, even inverting the two arrows no v-structured are created, therefore the three graphs that are generated by the reversions of the two arcs are equivalent and not distinguishable by any statistical test.

Instead, the arc from `RoadWorks` to `RoadConditions` cannot be reverted since doing this a v-structure is created. Hence, the arc could only be turned on condition of turn also the one between `Weather` and `RoadConditions`.

D-Separation

D-Separation tells when two variables are d-separated along a path (blocked), that means independent and when they are d-connected along a path (unblocked) or likely dependent. They are actually independent if they are d-separated along all possible paths. They are likely dependent if there is at least one unblocked path connecting them.

A path is blocked by a set of nodes if and only if the path contains a chain of nodes or a fork such that the middle node is in the set of nodes or if the path contains a collider such that the collision node and every descendant are not in the given set of nodes.

- `RushHour` and `RoadConditions`:
  Conditioning on one of the variables `Strike`, `RoadWorks` `Weather` or `Humor`, `RushHour` and `RoadConditions` are d-separated since the path from these two variables are all blocked. They are not actually independent since they are not d-separated along the paths that condition on variables `Danger`,`Speed`, `Accident` and `Delay` since in all this cases the path contains a collider in which the given is a collision node. In fact, all the variables d-connected are dependent in the real problem, for example `RushHour`, `RoadConditions` and `Danger`, while `RushHour`, `RoadConditions` and `Humor` are independent.

- `RushHour` and `Strike`:
  Conditioning on one of the variables `Danger`, `RoadWorks`, `Weather`, `Speed` or `Humor`, `RushHour` and `Strike` are d-separated since the path from these two variables are all blocked. Instead, they are not d-separated along the paths that condition on variables `Accident` and `Delay` since in both the cases the paths contain a collider in which the given is the collision node. Hence, they are not independent. In the real problem, in fact, the d-connected variables are dependent, for example `RushHour`, `Strike` and `Danger`, while `RushHour`, `Strike` and `RoadWorks` are independent.

- `Speed` and `Accident`:
  Are not d-separated given any of the variables of the domain.

- `RoadConditions` and `Strike`:
  In this case the same considerations made for the variables `RushHour` and `Strike` in example 2 apply.

- `Speed` and `Humor`:
  Conditioning on one of the variables `RoadConditions` or `Weather`, `Speed` and `Humor` are d-separated since the path from these two variables are all blocked. Instead, they are not d-separated along the paths that condition on all the other variables, so they are not independent. The d-connected variables, for example `Speed`, `Humor` and `Danger`, are dependent in the real problem, while `Speed`, `Humor` and `RoadConditions` are not dependent.

2 Conditional Probability Tables

The Conditional Probability Tables (CPTs) of the variables of the model, that show all possible inputs and outcomes with their associated probabilities, are filled sometimes using information retrieved from online survey, other times are estimated based on common sense. In case of the variable `Weather`, since his prior probability is difficult to estimate because it is dependent on different factors, such as the location, the probabilities correspond to a uniform distribution.

Weather	
Sun	Rain
0.5000	0.5000

Figure 2: Weather CPT

Strike	
True	False
0.1000	0.9000

Figure 3: Strike CPT

RushHour	
True	False
0.2000	0.8000

Figure 4: RushHour CPT

	Humor	
Weather	Good	Bad
Sun	0.8000	0.2000
Rain	0.3000	0.7000

Figure 5: Humor CPT

	RoadConditions	
Weather	Dry	Wet
Sun	0.7500	0.2500
Rain	0.4000	0.6000

Figure 6: RoadConditions CPT

	RoadWorks	
RoadConditions	True	False
Dry	0.1000	0.9000
Wet	0.8000	0.2000

Figure 7: RoadWorks CPT

			Speed	
RoadConditions	RushHour	RoadWorks	Slow	Fast
Dry	True	True	0.8500	0.1500
		False	0.8000	0.2000
	False	True	0.7500	0.2500
		False	0.1500	0.8500
Wet	True	True	0.9500	0.0500
		False	0.8000	0.2000
	False	True	0.9000	0.1000
		False	0.6000	0.4000

Figure 8: Speed CPT

			Danger	
RoadConditions	RushHour	Speed	Low	High
Dry	True	Slow	0.8500	0.1500
		Fast	0.2000	0.8000
	False	Slow	0.9500	0.0500
		Fast	0.3000	0.7000
Wet	True	Slow	0.4500	0.5500
		Fast	0.0500	0.9500
	False	Slow	0.5500	0.4500
		Fast	0.2000	0.8000

Figure 9: Danger CPT

			Accident	
Danger	Strike	Humor	Low	High
Low	True	Good	0.1500	0.8500
		Bad	0.7000	0.3000
	False	Good	0.0500	0.9500
		Bad	0.5500	0.4500
High	True	Good	0.8500	0.1500
		Bad	0.9500	0.0500
	False	Good	0.6000	0.4000
		Bad	0.8000	0.2000

Figure 10: Accident CPT

			Delay	
Speed	Strike	Accident	True	False
Slow	True	Low	0.7500	0.2500
		High	0.9800	0.0200
	False	Low	0.6500	0.3500
		High	0.9500	0.0500
Fast	True	Low	0.6000	0.4000
		High	0.9000	0.1000
	False	Low	0.4500	0.5500
		High	0.8500	0.1500

Figure 11: Delay CPT

3 Causal Inference

Choose one pair of variables. The pair must be made up of a variable X with at least one parent and another variable Y of the graph such that there is (at least) a causal path from X to Y. For the pair (X,Y) perform: • Calculate the causal effect of X on Y. • Identify possible confounders between X and Y. • Would it be practically possible in your specific problem to perform also a randomized controlled study to disentangle the causal effect between the variables from their correlation? • Compute the ACE of X on Y. Choose another pair of variable (X,Y) (it can be also the previous one) and: • Choose another variable C such that it is possible to calculate the c-specific effect of X on Y and calculate it. • Identify a minimal set of variables that must be measured in order to estimate the c-specific effect of X on Y. • Choose a function g and compute the effect of the conditional intervention of X=g(C) on Y. Choose another pair of variable (X,Y) (it can be also the previous one) and: • Identify possible mediating variables between X and Y and calculate the CDE of Y changing the value of X.

4 Simulation

Suppose that you can't measure some parents of variable X chosen in every point of "Causal Inference". Repeat the "Causal Inference" part of the exercise considering this new situation.

5 Comment on the Results

What kind of experience have you got with this model? E.g., is the causal model responding in a sensible way to your queries? What should be changed/modified to make it more realistic?