CS5011: Assignment 4 - Reasoning with Uncertainty - Bayesian Networks

Assignment: A4 - Assignment 4

Deadline: 22 December 2017

Weighting: 30% of module mark

Please note that MMS is the definitive source for deadline and credit details. You are expected to have read and understood all the information in this specification and any accompanying documents at least a week before the deadline. You must contact the lecturer regarding any queries well in advance of the deadline.

---

1 Objective

This practical aims to learn how to model and use Bayesian networks to reason with uncertainty applied to diagnostic systems.

2 Competencies

- General understanding of Bayesian networks
- Design, analyse and document different Bayesian networks to perform probabilistic reasoning, using a Bayesian network tool.

3 Practical Requirements

3.1 Introduction

Bayesian Belief networks (BNs) are a way of representing probabilistic relationships among a set of variables. They are often much more compact than the full joint probability distribution. They constitute an effective systematic way to represent uncertain representation of the world. Bayesian Networks are probabilistic graphical models representing a set of random variables and their conditional dependencies via directed acyclic graphs. Seminal research in the context of Bayesian Networks and causal reasoning, has led Judea Pearl[1] one of the first pioneer of BNs, to being awarded the prestigious Turing Award in 2011[2]. Bayesian Networks have important applications in diagnostic expert systems particularly for medical applications, investigation of faulty systems, and legal domains. In recent years, Bayesian Networks have also seen interesting applications in classification tasks, as basis for

[1] `http://amturing.acm.org/bib/pearl_2658896.cfm`
[2] `http://amturing.acm.org/alphabetical.cfm`

Bayesian Network Classifiers (See Chapter 20 of Russel & Norvig[3], CS5010 BN-L01). In this practical, you will gain familiarity with modelling and using Bayesian Networks with the help of a tool.

In the practical, we will use the AIspace Belief and Decision Networks tool (version 5.1.10, August 9th, 2016) available from `http://aispace.org/`. Download and save the *bayes.jar* file from the download page, then run it from terminal as:

```
java -jar bayes.jar
```

Other methods to run the tool can be found on the AIspace website. We will refer to this tool as the AIspace tool for simplicity.

3.2 Part 1 - Analysis/Use of BNs

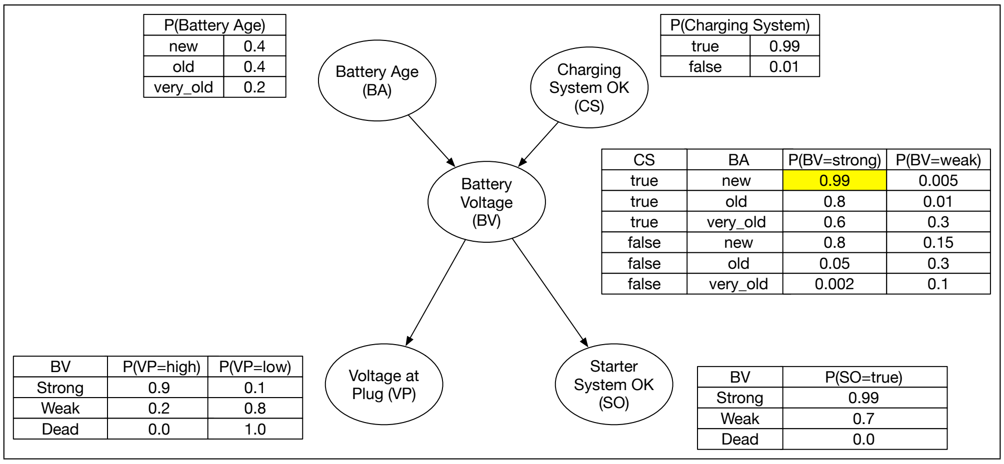
In this part of the practical, you will be investigating and analysing two given Bayesian Networks.

3.2.1 BN for Car Faults investigation

Consider the Bayesian Network presented in Figure 1. This example is a simplified version of the AIspace Car Starting Problem, to model the behaviour of a car based on the state of its parts. In this network we only focus on the car battery and related problems.
Tasks to be performed:

1. Construct the network using the AIspace tool, under the *create* view.
2. Perform the following queries, using the AIspace querying function under the *solve* view, and discuss your findings in the report, including observations on your expectations and the results obtained.
    (a) *Diagnostic*: The agent observes that the car starter system is not OK, what other characteristics would the car present?
    (b) *Diagnostic*: The agent observes that the voltage at plug is low and the car stater system is not OK, what other characteristics would the car present?
    (c) *Profiling*: Determine the general characteristics of a car and what issues it might have depending on its battery age.
    (d) *Predictive*: If the starter system and the charging system are both not ok, what is the likely situation of the voltage at plug?
3. Save the network as *prob1.xml* and submit it with your assignment.

[3]S. J. Russell and P. Norvig. Artificial Intelligence: A Modern Approach. Pearson Education, 3 ed., 2010.



P(Battery Age)	
new	0.4
old	0.4
very_old	0.2

P(Charging System)	
true	0.99
false	0.01

CS	BA	P(BV=strong)	P(BV=weak)
true	new	0.99	0.005
true	old	0.8	0.01
true	very_old	0.6	0.3
false	new	0.8	0.15
false	old	0.05	0.3
false	very_old	0.002	0.1

BV	P(VP=high)	P(VP=low)
Strong	0.9	0.1
Weak	0.2	0.8
Dead	0.0	1.0

BV	P(SO=true)
Strong	0.99
Weak	0.7
Dead	0.0

Figure 1: BN for car faults Part 1.2

3.3 BN for medical diagnosis

In the AIspace tool, load the simple diagnostic example, by selecting File→Load Sample Problem→Simple Diagnostic Example [4]. This example is about a knowledge base for an autonomous medical diagnostic assistant. The BN shown in Figure 2 should be displayed.

> The agent can observe coughing, wheezing, and fever and can ask whether the patient smokes. There are variables that are useful to predict the outcomes of patients: Bronchitis or Influenza. The variables directly depend on are the following: whether patients wheeze depends on whether they have bronchitis; whether they cough depends on whether they have bronchitis; whether patients have bronchitis depends on whether they have influenza and whether they smoke. Whether they have fever depends on whether they have influenza.

Tasks to be performed:

1. Use the Query function of the tool to perform the following investigation:
   (a) *Diagnostic*: The agent observes that the patient is wheezing, how likely is that the patient smokes?
   (b) *Predictive*: The agent observes that the patient is a smoker, how likely is that the patient will develop bronchitis? And that they will have sore throat?
   (c) *Predictive*: The agent observes that the patient is a smoker, and has influenza, how likely is that the patient will develop sore throat?
   (d) *Predictive*: The agent observes that the patient is a not smoker, and has influenza, how likely is that the patient will be coughing?
   (e) *Intercausal*: The agent observes that the patient is not a smoker and, has got bronchitis. How likely is that they have influenza?

[4] This is example 6.14, shown in Figure 6.3 of Poole and Mackworth, Artificial Intelligence: foundations of computational agents, 2010.

(f) *Profiling*: Determine the characteristics of patients that smoke and compare those against patients that do not smoke.

For each of the above queries, explain your expectations and findings on the computed probabilities.

2. We can see that the flu has a further symptom, fever, and we now assume that this can be measured by a thermometer *Therm*, but the readings may be somewhat unreliable. If there is fever, Therm will indicate fever 90% of the times, if there is no fever, *Therm* may still indicate Fever 5% of the times. Add this node to the network and perform the following *predictive* query. If the patient has influenza, how likely is that the thermometer will indicate fever? Discuss the result in your report.

3. Save the network as *prob2.xml* and submit it with your assignment.

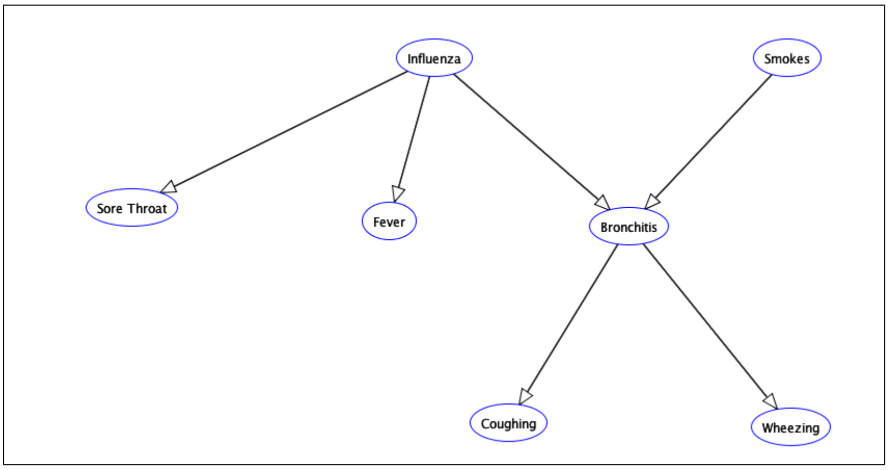


Figure 2: BN for simple diagnostic example Part 1.1

3.4 Part 2 - Design of BNs

In this part of the practical, you will be designing a neural network for the following scenario.

Suppose you are designing an alert system for traffic congestion that would work in coordination with a real route planner (eg. Google Maps). Assume that we want to model the following alert system to predict the congestion at the BlueRainbow bridge (BRB for short). The system relies on the following sensors and data gathered from information systems and that are related to the traffic congestion.

1. A visibility sensor indicates that there is either High or Low visibility in BRB depending on the weather (sunny, overcast, raining). The sensor is pretty good at detecting visibility parameters but may give the wrong value with 1% probability.

2. A roadwork information system, which indicates whether there are or not planned roadworks in BRB. Normally, there is a 10% probability that there is roadwork in BRB, and furthermore, that information is 3% of the times out of date.

3. Traffic data analytics regarding the time of day indicate whether in BRB it is peak time (between 6-10 and 4-8 for a total of 8 hours) or off peak time (total of 16 hours in a day).

4. The council information system also provides information on whether it is a working day or a holiday day (assume 90 days holidays in 360 days). Furthermore, national statistics show that on holiday periods 60% of people take their car, while during non-holiday periods there is a probability of 70% that people will take the car.

5. An alert will be triggered (and shown in the route planner) if we predict traffic congestion with 80% probability.

6. A piezoelectric sensor, for vehicle count, is placed at the entrance of BRB, and indicates that there is either a High or Low traffic flow. This sensor has a bad fault tolerance and may give the wrong value with 10% probability. While not used for triggering alerts, this can be useful for checking the predictions.

Assume that all the above factors are independent and that we want to know about the traffic congestion situation right now, without considering how the situation evolves in time.
Tasks to be performed:

1. Construct a Bayesian Network to represent and draw inferences in this problem:
   - Decide what your domain variables are and what the relationships are between the domain variables
   - Then add the conditional probabilities for nodes that have parents, and the prior probabilities for nodes without parents.
   - Use the information given to fill these probabilities, and for those probabilities that haven't been given explicitly choose values that seem reasonable and explain why in your report.
   - Construct the Bayesian network in the tool, save it as *prob3.xml* and submit it with your assignment.

2. In the real world, the factors given to predict the congestion are not necessarily independent. For example, holiday periods may influence peak time traffic information and/or planned roadworks. Another example may be the case in which all of information those systems are powered by a poor quality power supply, the information may then be faulty or out of date. Adapt your bayesian network to have some degree of correlation among variables either through new or existing events. Construct the Bayesian network in the tool, save it as *prob4.xml* and submit it with your assignment.

3. For both Bayesian networks, use the Query function to investigate changes of probability distribution giving examples of at least 3 *predictive*, and 2 *diagnostic* tasks. Present these examples and summarise your findings in your report.

3.5 Part 3 - Extensions

You may wish to consider one or more of the extensions listed below:

- Define your own problem involving reasoning with evidence and uncertainty. Write down a text description of the problem, then model it using a Bayesian network in the AIspace tool. Make the problem sufficiently complex such that there are at least 5 nodes and some degree of correlation among variables. Save and submit your network, and use the Query function to investigate changes of probability distribution giving examples of *predictive* and *diagnostic* tasks. Present these examples and summarise your findings in your report.

- Use any of the networks proposed for this assignment and implement it using a java library for Bayesian Networks (Encog for Java[5] may be used for this step). Then develop an application for an expert agent assistant. The system should include a text-based interface that can be used to query the BN as an expert system. The user should be able to receive support for *diagnostic* or *predictive* tasks and the agent should show the results in a user friendly way.

- Find out about Decision Networks in Chapter 16.5 of Russel & Norvig[6]. Modify the network developed in Part 2 to decide on whether it is worth to take the BRB route by including some utility values. Save and submit your network, giving some explanation on how the network could be used for this task.

- Suggest your own extension, a significant extension should consider a design of a bayesian network or the use of a bayesian network.

4 Submission and Running

For part 1 and part 2, you are required to submit XML files created with the BN tool from AISpaces. Please name your files as

```
 Part 1) prob1.xml prob2.xml
Part 2) prob3.xml prob4.xml
```

For part 3, depending on your choices, if a new network is proposed, please submit an .xml file and indicate the name of the network in your report. If a system is implemented, the program must be written in Java and your implementation must compile and run on the School lab Machines. A jar file should be submitted in this case that runs with the command:

```
java -jar prob.jar [any param]
```

[5] http://www.heatonresearch.com/encog/

[6] S. J. Russell and P. Norvig. Artificial Intelligence: A Modern Approach. Pearson Education, 3 ed., 2010.

5 Report

You are required to submit a report describing your submission, with a limit of 5400 words excluding references. The report should include:

1. A list of the parts implemented and any extension attempted
2. If any of the functionalities is only partially working, ensure that this is discussed in your report
3. Literature review: a short literature survey on Bayesian networks and their current applications.
4. Depending on the tasks performed, include:
    (a) Part 1:
    - Design: procedure for investigation
    - Evaluation: results and critical analysis of the given tasks

    (b) Part 2:
    - Design: procedure for network design
    - Testing: examples and testing of the networks
    - Evaluation: results and critical analysis of the given tasks

    (c) Part 3 - Depending on the extension attempted, points similar to Part 2 should be included or if a system has been developed include the usual Design, Examples and Testing, and Evaluation of the system:
    - Design: A description and justification for the mechanisms you have implemented as part of your solution.
    - Examples and Testing: Examples of the main functionalities and your approach to testing the system.
    - Evaluation: A critical analysis of the functionalities of your system
5. A general evaluation section discussing your achievements and including points that could be improved.
6. Running: Include clear instructions on how to run your system
7. Bibliography: List all the references you cite in your literature review and elsewhere in your report and code.

Please include a word count in your report.

6 Deliverables

A single ZIP file must be submitted electronically via MMS by the deadline. Submissions in any other format will be rejected.

Your ZIP file should contain:

1. A PDF report as discussed in Section 5
2. Depending on the tasks achieved, a set of BNs in XML format and any jar of code implemented in Part 3, as discussed in Section 4
3. The source code of your implementation of Part 3 if any, containing any non-standard libraries

7 Assessment Criteria

Marking will follow the guidelines given in the school student handbook (see link in the next section).

The following issues will be considered:

- Adherence to the requirements
- Quality of the solution provided
- Examples
- Insights and analysis demonstrated in the report
- Extensions activities completed, if any

Some guidelines are as follows. For a mark up to the band 11-13, Part 1 should be performed. For a mark up to 17 you must complete Part 1 and Part 2. To obtain marks above 17, in addition to Part 1 and 2, at least one significant extension should be included as those suggested in Part 3. All parts are to be accompanied by a good insightful report covering all points and good design and implementation quality.

8 Policies and Guidelines

8.1 Marking

See the standard mark descriptors in the School Student Handbook

`https://info.cs.st-andrews.ac.uk/student-handbook/learning-teaching/feedback.html#Mark_-Descriptors`

8.2 Lateness Penalty

The standard penalty for late submission applies (Scheme B: 1 mark per 8 hour period, or part thereof):

`https://info.cs.st-andrews.ac.uk/student-handbook/learning-teaching/assessment.html#latenesspenalties`

8.3 Good Academic Practice

The University policy on Good Academic Practice applies:

`https://www.st-andrews.ac.uk/students/rules/academicpractice/`

Alice Toniolo
(a.toniolo@st-andrews.ac.uk)
November 28, 2017