RADBOUD UNIVERSITY

How to Evaluate Network Models Used in Clinical Decision Making

Author:
Patrick Tan
s0032573

First supervisor/assessor:
Prof.dr. Peter Lucas
peterl@cs.ru.nl

Second assessor:
Dr. Patrick van Bommel
p.vanbommel@cs.ru.nl

July 11, 2016

Abstract

Due to the rising complexity of medical decision problems there is an increasing need for decision support. In an effort to support physicians in their medical decision making, network systems have been developed which can weigh diagnostic and patient information and recommend treatment options that also offer a favourable prognosis. In order to promote the use of these network systems it is desirable that there is a way to show the practical value of these network models. The context of the research described in this thesis is clinical treatment selection. There network models are commonly used.

I have compared several methods from different articles and books that have something to say about the development and evaluation of network models and have arrived at a methodology to evaluate network models used for clinical treatment selection. The way in which each network model should be evaluated may differ per model due to the decision problem it is expected to handle. Several difficulties may need to be overcome during evaluation. However, it appears that it is possible to evaluate these network models to form an informed opinion regarding their functioning.

Contents

Chapter 1

Introduction

Making decisions has always been a part of life as we know it. Though some might argue that because we live in a deterministic universe our actions are predetermined the actual decisions regarding those actions still need to be made on a conscious level.

Predetermined or not, most organisms are wired in such a way to make decisions that increase their chance of survival. Simple organisms may simply react to their sensory input, more complex organisms can tap into their memory and integrate previous experiences into making decisions. Even more complex, higher organisms if you will have the ability to reason and may make decisions based memories and beliefs not directly linked with the decision at hand.

We as humans have this ability to reason. We use it on a daily basis and in many aspects of our lives. What we should eat, wear and read. We make many decisions during the day. Of course, being humans, we have professions, which often entail making decisions of a higher complexity than what we would like to have for dinner.

One group of professionals that deals with particularly important decisions would be those involved in the field of medicine. Medical practitioners need to extract data from their patients and use this to ultimately come up with the decision most beneficial to the patient and to which the patient can give informed consent to [6, 25, 29]

As medical science grew more advanced, so too did the decision making problems involving treatment selection become more complex, not in the least because medical practitioners often have to deal with uncertainty and have to work with rather imperfect information. Medical practitioners are also under increasing scrutiny as patients gain more access to information

themselves. Also, having to deal with qualia such as discomfort, aesthetics, perhaps even happiness, means that the decisions are also partially made subjectively which makes it harder to judge the quality of a decision. Making a decision that will lead to the most desirable outcome to the best of one's ability therefore requires a combination of solid reasoning an intuitive understanding of what the patient prefers. Intuition is still beyond computers, but a clinician may be able to get some support regarding the rational part of the decisions made in the form of a decision support system [27]. Network models are one form of support system that can fulfil this function [22]. A well-made network model may make help see the clinician see structure in the decision problem and also calculate, given the evidence which course of action would lead to the optimal outcome [9, 16].

These decision making support systems have been under development for several decades now. It seems almost inevitable that they are going to see widespread implementation. Still medical practitioners are sceptical about these models as they should be about everything that concerns the welfare of their patients. They need to be reassured that these models fulfil their function and that using them will beneficially affect healthcare [12].

In order to promote the use of these network models for clinical decision making it is important that these models can show their worth. They need to be properly evaluated. This might not be an easy task. We have already established that some variables are subjective in nature and evaluation is a challenge in itself. Still for the sake of improving healthcare we must try, and that is why my research question is "How to evaluate network models used in clinical decision making?".

First I shall briefly explain the basics of decision theory, probability theory and graph theory as these are necessary to get a grasp on the problem at hand. I will then explain which models are most commonly used and how their results are expressed and measured. Then I shall proceed to explain how clinical treatment selection is performed and how a model might support this. I will then describe the results of my research by explaining at which points during the lifespan of a network model one might evaluate it and how this is done. Lastly I will explain my conclusion.

Chapter 2

Preliminaries

Before we can describe something new we must explain what is already known. We shall limit ourselves to what is relevant within the scope of this project. We shall describe probability theory, graph theory, Bayesian networks, statistic decision trees and influence diagrams.

2.1 Decision theory

2.1.1 What is decision theory?

Decision theory is defined as the analysis of the behaviour of an individual facing uncertainty [13]. It is used in many fields but the way it is used in medical science is of particular interest to us. More specifically how a clinician decides upon which treatment to recommend to a patient. A decision-making task is one where a decision or choice is made. Usually the result is based on which decision has been made and some results are more desirable than others. A rationally made decision therefore is based on what the outcome will be and the decision maker chooses the option which leads to the most desired outcome. In situations where the outcome of every possible decision is known and it is clear which of these is the best and the decision-making is trivial. Such situations are not analysed by decision theory.

2.1.2 Utility

One of the challenges of decision theory is to define what is the best outcome. One way to do this is to assign each outcome a *utility score* [25,28,29]. If the

desirability of the outcome is based on one variable then so will the utility score be.

If our goal is to find as much fruit as possible then the utility score will solely be based on the amount of fruit found. When multiple variables determine how desirable the outcome is utility can be assigned to each variable and the final utility score will be the sum of these individual utility scores.

Suppose someone tries to find as much fruit as possible, but only apples, pears and oranges. The total utility score will be based on the sum of the apples, pears and oranges found. Not every variable needs to be assigned the same utility score. If for example an apple is more desirable than a pear, and it is known exactly by how much then the utility score can express that by assigning more utility for every apple found compared to how much utility is assigned for every pear found.

If something is considered undesirable it may be assigned a negative utility score. Some other complicating factors in decision making would be the interaction of decision makers and situations where the time at which an action is taken is also of influence, also being referred to as the temporal factor. In medical decision making variables that could be used to calculate the utility score could be

- Treatment cost in euros

- Discomfort inflicted upon the patient

- Immediate chance of survival for the patient

- Chance of a full recovery

- Length of the recovery time

- Life expectancy of the patient

- Resulting quality of life for the patient

- Availability of specialists for certain treatments

Of course some of these variables are not easily quantified such as discomfort inflicted upon patient and resulting quality of life. This means that assigning a utility score for these variables is a subjective matter. Also, not every patient finds each of these variables equally as important which means weighing the variables against each other to calculate the utility score is a subjective matter [8].

6

In addition, variables such as treatment cost may be quantifiable, but can these costs be justified? Is it exactly twice as undesirable to spend 200 euro as it would be to spend 100 euro? And how do you properly anticipate on future cost [20]?

In short one should always be wary that utility is not as clean cut a parameter as one might think.

2.2   Probability theory

Probability theory is a branch of mathematics that analyses phenomena perceived as random. Phenomena of which the outcomes are considered to be uncertain [9, 16–18, 24, 25].

A process of which the outcome is not known in advance is called an *experiment*. The set of all possible outcomes is called the *sample space S* of the experiment.

A *probability distribution P* is a function that associates outcomes with the probability of their occurring. A subset of the sample space is called an *event E*. An event is considered to be true if the outcome is an element of the event, otherwise it is considered to be false which is expressed as $\bar{E}$. Every event has a probability to occur, in other words $E \subseteq S$. This is expressed as $P(E)$. $\bar{E} = S \setminus E$ which is the complement of event E.

If we look at a single roll of a single six sided die the sample space would be expressed as follows.

Sample space $S = \{1, 2, 3, 4, 5, 6\}$.

An example of an event would be that an even value has been rolled.

$E = \{2, 4, 6\} \subseteq S$.

The probability of event E is written as $P(E) = 0.5$. In this case the probability of $\bar{E} = S \setminus E = \{1, 2, 3, 4, 5, 6\} - \{2, 4, 6\} = \{1, 3, 5\}$ The probability of this event is also 0.5. Not surprising because if the likelihood of an event is 0.5 then the likelihood of its complement is 1 - 0.5 = 0.5.

In 1933 Kolmogorov published a book titled "Foundations of the Theory of Probability". Where he introduces the three axioms [17] discussed in the next section.

7

2.2.1 Kolmogorov's three axioms

1. Kolmogorov's first axiom for probability theory

   $$P(S) = 1$$

   The set of all possible outcomes is the sample space and therefore the probability of the sample space is 1.

2. Kolmogorov's second axiom for probability theory

   $$\forall E \subseteq S : P(E) \geq 0$$

   Any event E must have a probability that is not negative.

3. Kolmogorov's third axiom for probability theory

   $$E_1 \subseteq S \wedge E_2 \subseteq S \wedge E_1 \cap E_2 = \emptyset \to P(E_1 \cap E_2) = P(E_1) + P(E_2).$$

   When two events are disjoint the probability of the union of those two events is the sum of the probability of each individual event.

Axiom 1   Axiom 2   Axiom 3
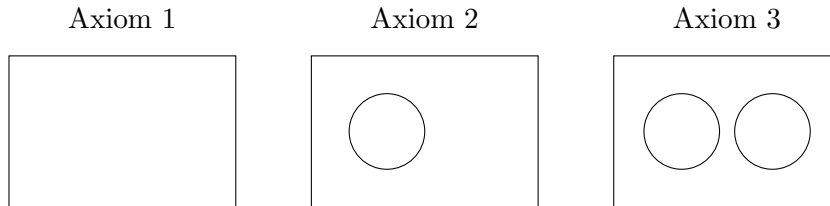
Figure 2.1: Kolmogorov's 3 axioms

This rectangle represents a 2 dimensional sample space. We shall consider the surface of this rectangle 1.

A circle represents an event and its surface represents the space it takes in the sample space

When two events are disjoint it can be seen that the probability of either event taking place is the sum of their surface areas.

2.2.2 Conditional probability

Sometimes you will want to know the probability distribution of something after learning that a certain event is true. For example, what can you say

about the odds of someone ordering a cup of coffee given that he has ordered a bagel?

Formally this is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

with $P(B) > 0$.
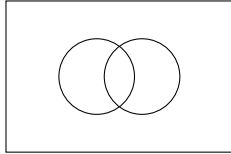
Figure 2.2: Conditional probability

In the figure below we could use conditional probability to calculate the probability of a point in the sample space being in the right circle when we already know it to be within the left circle. This probability is represented by the area where the circles overlap. What it essentially does is limit the space of outcomes to one of the circles and check the inside of that circle for the odds of a point being inside the second circle as well.

2.2.3   Chain rule

The probability of a combination of events can be expressed as the probability of the third given the second given the first. This holds for any number of events and the result will be the same regardless of the order of events. It is written as

$$P(A_1 \cap \cdots \cap A_k) = P(A_1)P(A_2|A_1) \cdots P(A_k|A_1 \cap \cdots \cap A_{k-1})$$
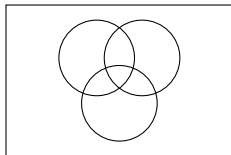

Figure 2.3: Chainrule

As the figure shows we can use the chain rule to calculate the probability of a point in the sample space being within all three circles. The order in

which we check the circles will not affect how large the surface area of where the three circles overlap is.

2.2.4 Bayes' theorem

Bayes theorem is often used to calculate the probability distributions of variables related to an event given the observed outcome. It reads as follows.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

It means that the probability of an event given some evidence is the probability of the evidence given the event times the probability of the event without regard to the evidence divided by the probability of the evidence without regard of the event.

For example, suppose I have two six sided dice. I want to roll higher than 9. The sample space would have 36 possible outcomes of which 6 would make my roll higher than 9. My prior belief or prior is therefore that my chance of rolling higher than 9 is $\frac{6}{36}$. We shall call this roll of higher than 9 event A. If I roll the first die and it is a 5 then I will have gathered some evidence. We can use Bayes theorem now to calculate what the odds are now.

The probability of rolling higher than 9 given that the first die came up with a 5 would be the probability of rolling a 5 given that we did roll higher than 9 times the probability of rolling higher than 9 divided by the probability of rolling a 5 with the first die.

$$P(\text{higher than 9}|\text{first die rolled 5})$$

$$= \frac{P(\text{first die rolled 5}|\text{higher than 9})P(\text{higher than 9})}{P(\text{first die rolled 5})}$$

$$= \frac{\frac{2}{6} * \frac{6}{36}}{\frac{1}{6}} = \frac{1}{3}$$

We know this to be correct because if the first die roll is a 5 then only a 5 or a 6 on the second die roll will yield a result higher than 9, which has a probability of $\frac{1}{3}$. This is known as the posterior probability or *posterior* for short. It is the *prior* with its beliefs updated given the evidence.

No evidence			Updated once			Updated twice	
Monday	1/7		Monday	0		Monday	0
Tuesday	1/7		Tuesday	1/6		Tuesday	1
Wednesday	1/7		Wednesday	1/6		Wednesday	0
Thursday	1/7		Thursday	1/6		Thursday	0
Friday	1/7		Friday	1/6		Friday	0
Saturday	1/7		Saturday	1/6		Saturday	0
Sunday	1/7		Sunday	1/6		Sunday	0

2.2.5 Bayesian philosophy

Bayesian philosophy is based on Bayes' theorem and follows the belief that if a hypothesis has been formulated with a probability or belief assigned to it and some evidence comes to light then the probability or belief should be updated to reflect that evidence [18].

For example, if I state I will be going to my favourite restaurant sometime next week and say nothing else and we assume I will only be going once that week the probability distribution would show that each day has a probability of 1 in 7 to be the day I visit my favourite restaurant. Now if we add a piece of evidence, namely that the particular restaurant is closed on Mondays then the beliefs must be updated accordingly. It would show a probability of 0 for Monday and 1/6 for the other days. If I were to add additional evidence that I will be going with a friend of mine who can only go on Tuesdays then the beliefs will again be updated to reflect this.

The task of belief updating is also known as *probabilistic inference.*

2.2.6 Conditioning

Conditioning is used to reduce the number of calculations needed to determine the probability of an event by establishing that a certain condition has been met. If the value of a variable is known which influences the outcome of an event then we can use this evidence to do a calculation which uses the known value. A simple example would be the situation in which we were to roll two six sided dice and we want to know the odds of throwing more than nine as explained previously. When none of the dice have been thrown the odds would be calculated by calculating the possibility of throwing nine or more looking at the distribution of each die individually. This would involve

calculating the odds for every individual value of the first die to be thrown which is 1/6th for each value. But then for each of those possible values it has to be calculated separately what the odds are that the value of the first die will yield nine or more when added to the value of the second die.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

If we already threw one die we no longer need quite as many calculations. Suppose a 5 has been thrown. We basically check the equivalent of one row of the table. Like this.

	1	2	3	4	5	6
5	6	7	8	9	10	11

Thus reducing the number of outcomes that have to be checked.

If we know B is dependent on A and we have evidence regarding A we can condition this by calculating $P(B|A)P(A)$ and $P(B|\neg A)P(\neg A)$, thus limiting the sample space we need to look into and making our calculations less costly.

2.2.7 Marginalization

Marginalization is the elimination of variables out of the joint probability distribution P.

Given two random variables X and Y with a joint probability distribution $P(X, Y)$:

	Y	$\neg Y$
X	0.30	0.15
$\neg X$	0.10	0.35

It is possible to calculate the marginal distribution of X by averaging the information about Y : $P(X) = P(X, Y) + P(X, \neg Y)$;

X	0.45
$\neg X$	0.55

2.3   Graph theory

Graph theory is the study of graphs. A graph is a mathematical structure that models the pairwise relations between objects and are typically used to graphically illustrate mathematical structures allowing for a more intuitive interpretation.

Elements of graphs consist of *nodes*, also called *vertices*, and *edges*, also called *arcs*. For example graph $G = (V, E)$ where V is the collection of vertices or nodes and E is the collection of edges. An edge is related to two vertices, thus signifying a relationship. In a graphical representation a node is typically a circle and an edge would be a line connecting two circles.
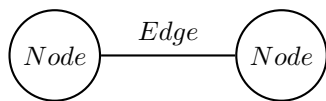

Figure 2.4: A simple graph

An edge may connect a node with itself, signifying that the node has a relation with itself somehow. Nodes may also be connected by multiple edges, signifying multiple relations between the nodes in question. An example for this would be in a graph that shows distances between cities and there happen to be multiple roads between two cities. This is considered a multigraph.

There are different types of graphs but most significant would be the difference between *directed* and *undirected* graphs. In a directed graph an edge between vertex V_a and V_b would only imply the relation $V_a \rightarrow V_b$ or $V_b \rightarrow V_a$. The relation goes only one way.

In an undirected graph an edge between vertex V_a and V_b implies both $V_a \rightarrow V_b$ and $V_b \rightarrow V_a$, sometimes expressed as $V_a \leftrightarrow V_b$, implying that the relation goes both ways, which means it is a reflexive relation.
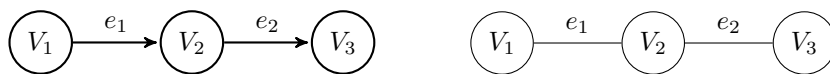
Figure 2.5: A directed graph and an undirected graph

A graph is considered cyclic if it is possible to make a path that visits the same node multiple times.
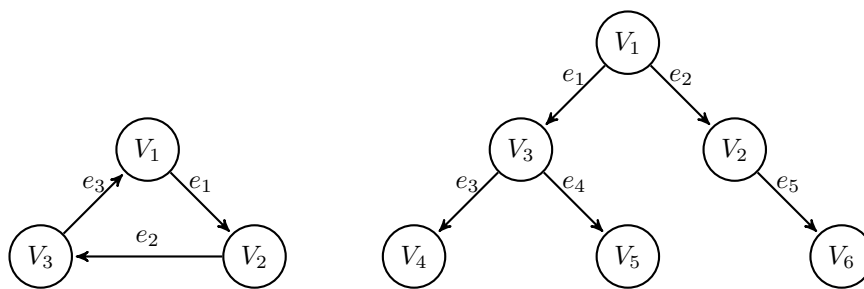
Figure 2.6: A cyclic graph and a tree

2.3.1 Trees

Formally a *tree* T is a graph $T = (V(T), A(T))$, which is a directed acyclic graph in which $V(T)$ is a set of vertices and $A(G)$ is a set of (directed) edges. Furthermore, there is one vertex, called the *root* $r \in V(T)$, that has no incoming edges, and there is always one directed path from the root r to any other vertex in $V(T)$. In this example V_1 is the root. $e_2 = (V_1, V_2) = V_1 \rightarrow V_2$

Chapter 3

Models used in clinical decision support

Combining knowledge from decision theory, probability theory, graph theory and medical science we can create network models that can function as clinical support systems. These models can use the evidence gained from diagnostics and give insight about what the likelihood of certain outcomes should a certain treatment be selected. I shall describe the most commonly used network models used for this purpose.

3.1 Bayesian networks

3.1.1 Formal definition of Bayesian networks

Formally a Bayesian network $\mathcal{B} = (G, P)$ is a pair, where $G = (V(G), A(G))$ is a directed acyclic graph in which $V(G)$ is a set of vertices and $A(G)$ is a set of edges and

$$P : \wp(V(G)) \to [0, 1]$$

is a joint probability distribution such that

$$P(V(G)) = \prod_{i=1}^{n} P(V_i | \pi_G(V_i))$$

$\pi_G(V_i)$ denotes the set of immediate ancestors of V_i in G.

Meaning that this joint probability distribution can be written as a product of the individual density functions, conditional on their immediate ancestors or parents.
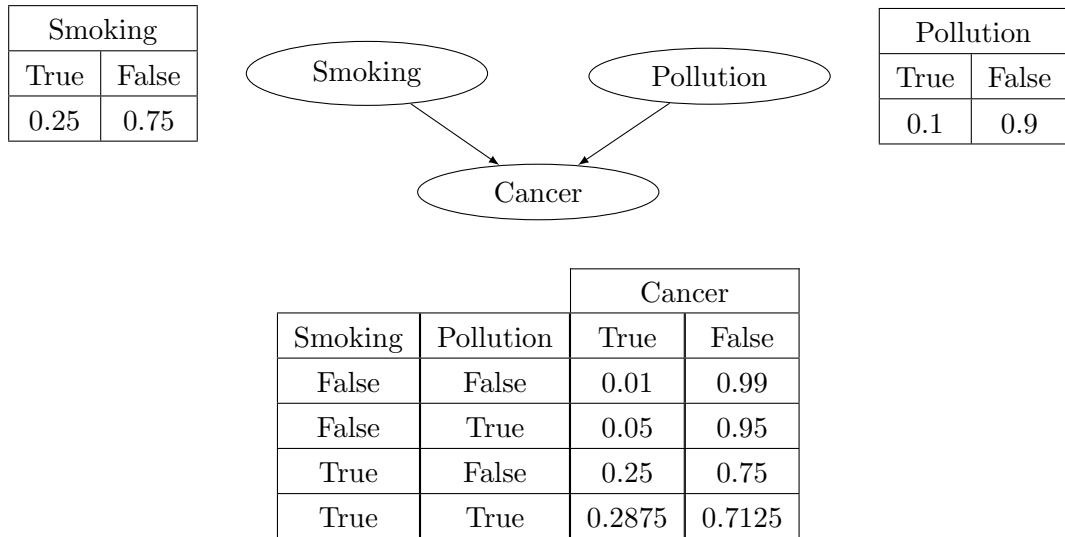
Smoking	
True	False
0.25	0.75

Pollution	
True	False
0.1	0.9

		Cancer	
Smoking	Pollution	True	False
False	False	0.01	0.99
False	True	0.05	0.95
True	False	0.25	0.75
True	True	0.2875	0.7125

Figure 3.1: A Bayesian network

3.1.2   A simple Bayesian network explained

This rather simple Bayesian network is used to shows us the relations between a form of cancer, smoking and pollution. According to this model a person that does not smoke nor is subjected to pollution might develop cancer. Smoking and pollution both increase the chance of developing cancer, as does a combination of the two. Smoking and pollution have cancer as their common effect. Notice that the likelihood of developing cancer when someone either smokes or is subjected to pollution does not specify whether it was the smoking or pollution that caused the cancer to develop or that it would have occurred without these factors.

The likelihood of developing cancer if the subject both smokes and is subjected to pollution is 0.2875, which can simply be looked up in the figure by looking up the field for which S, P and C are true. In short $P(C|S,P) = 0.2875$.

It is possible to calculate $P(C)$ using marginalisation. The calculation would look as follows.

$$
\begin{aligned}
P(C) = {} & P(C|S,P)P(S)P(P) \\
& + P(C|\bar{S},P)P(\bar{S})P(P) \\
& + P(C|S,\bar{P})P(S)P(\bar{P}) \\
& + P(C|\bar{S},\bar{P})P(\bar{S})P(\bar{P})
\end{aligned}
\tag{3.1}
$$

If we fill in the equation we get the following.

$$P(C) = 0.2875 * 0.25 * 0.1$$
$$+0.05 * 0.75 * 0.1$$
$$+0.25 * 0.75 * 0.9$$
$$+0.01 * 0.75 * 0.9)$$

(3.2)

$$P(C) = 0.0071875 + 0.00375 + 0.16875 + 0.00675 = 0.1864375$$

Therefore the table would look like this.

Cancer	
True	False
0.1864375	0.8135625

If we were to input some evidence, for example that we know a subject smokes then the probability distribution of pollution would remain unaltered because there is no path from smoking to pollution and vice versa. The likelihood of cancer would be affected. If we marginalize it since pollution is still unknown we would get the following calculation.

$$P(C|S) = P(C|S, P)P(P) + P(C|S, \bar{P})P(\bar{P}) = 0.1*0.2875 + 0.9*0.25 = 0.25375$$

Should we know for a fact that a subject has cancer then we can use inference to update the probability distributions of smoking and pollution. First we would have to normalize the probability distribution for cancer by multiplying each of the eight scenario's with the probability of it occurring. Below is a table which shows the likelihood of each combination of smoking and pollution as well as the normalized likelihood of each outcome.

			Cancer		Normalized	
likelihood	Smoking	Pollution	True	False	True	False
0.75 * 0.9 = 0.675	False	False	0.01	0.99	0.00675	0.66825
0.75 * 0.1 = 0.075	False	True	0.05	0.95	0.00375	0.07125
0.25 * 0.9 = 0.225	True	False	0.25	0.75	0.05625	0.16875
0.25 * 0.1 = 0.025	True	True	0.2875	0.7125	0.0071875	0.0178125

By using the presence of cancer as evidence we limit the sample space to a smaller sample space. This means we have to disregard all cases in which there is no cancer and normalize the results.

0.00675
0.00375
0.05625
0.0071875

becomes

0.0913
0.0507
0.761
0.0972

This would yield the following updated distributions for smoking and pollution. If we marginalize and normalize for both respectively.

Smoke	
True	False
0.8582	0.1418

Pollution	
True	False
0.1479	0.8521

3.1.3 Features of Bayesian networks

A Bayesian network is typically used to represent dependencies among variables. Each node in a Bayesian network corresponds to a variable and contains a probability distribution for the values of that variable. Edges signify a causal relationship between to variables and as such multiple edges leading to one vertex would signify a common effect relation while multiple edges emanating from a single vertex would signify a common cause relationship.

Distinguishing features of Bayesian networks are the following.

Inferring unobserved variables

When a model has been completed and some evidence is gathered regarding at least some of the variables the knowledge of unobserved variables can also be updated through inference. As demonstrated previously whenever evidence is gathered it is possible to update the probability distributions of the nodes directly connected to it by limiting the sample space, normalizing it and then marginalize. It is then possible to update the nodes connected to the nodes that were connected to the first node and so on. This phenomenon is called belief propagation and is one of the most important features of Bayesian networks [9, 14, 16, 18].

It is this feature that would allow a clinician to input the data he or has gathered regarding his patient, usually patient history and diagnostic information, and the model would show which outcome is the most likely given a certain treatment.

Parameter learning

If evidence has been gathered it can be checked how likely this evidence is to occur given the probability distributions in the Bayesian network. Using parameter learning it is possible to change the probability distributions in such a way that the likelihood that this evidence occurs is maximized. This is usually done by implementing a machine learning algorithm that uses data to train the network [26]. With this feature the probability distributions of a Bayesian network can be made to fit the data of a clinicians patient population better. The more data that is used for parameter learning, the more general the Bayesian network would become [9, 16, 18].

Structure learning

The structure of a Bayesian network can be designed using the knowledge of a domain expert [4]. An alternative way to find the structure is to use machine learning on data gathered to find the structure of the network [11,19]. The algorithm would try to find if certain variables show correlation and if so it would build the network on these correlations. Though not particularly useful to create Bayesian networks that give a good overview of the structure of the problem at hand this feature can yield Bayesian networks that function decently as classifiers.

3.1.4   Known disadvantages and limitations

Circular causality

Because a Bayesian network is acyclic it is not recommended to try to model circular causality in a Bayesian network. In such situations another type of model is recommended [23].

Subjectivity

When a Bayesian network is designed by a domain expert it will only be as good as the domain expert designing it. It might also be that another domain expert does not agree with the assumptions made while designing the Bayesian network such as the probability distributions and therefore would be sceptical about it functioning [25].

3.2   Decision trees

3.2.1   Formal definition of decision trees

Formally a decision tree T is a tree with a set of vertices $V(T) = C(T) \cup D(T) \cup U(T)$, which are mutually disjoint. $C(T)D$ is the set of chance vertices each of which has a probability distribution assigned to it, $D(T)$ is the set of decision vertices and $U(T)$ is the set of utility vertices each of which has a utility score assigned to it.

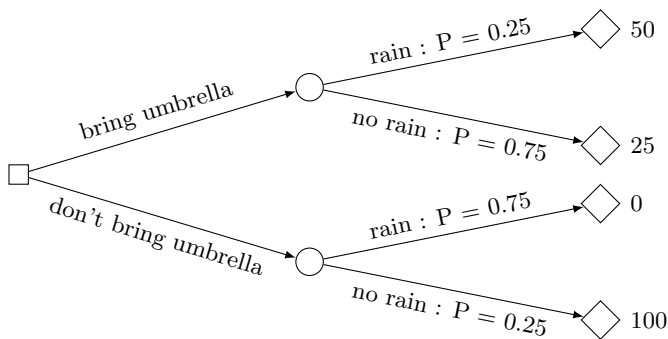### 3.2.2   Explanation of a simple decision tree

Figure 3.2: A decision tree

At the root of this tree there is a decision node. There are two options to choose from, bring an umbrella or don't bring an umbrella. The chance nodes that are connected to the decision node indicate that when one decides to bring an umbrella the probability of rain is 0.25 and when one decides not to bring an umbrella the chance of rain somehow is much higher at 0.75. Both chance nodes have two edges coming out of them to represent the possibility of rain and no rain. Each possible outcome has a utility score assigned to it. In this case the scenario with the highest utility score is considered the best outcome. We can now check what the expected utility is for each of the choices. If one choses to bring an umbrella it is 0.25 * 50 + 0.75 * 25 = 31.25 and for the option of not bringing an umbrella it is 0.75 * 0 + 0.25 * 100 = 25. Bringing an umbrella has the highest expected utility value, therefore according to the model one should bring an umbrella.

3.2.3   Features of decision trees

Decision trees are directed acyclic graphs which have three different kinds
of nodes.

- Decision nodes Usually represented by a rectangular box. Decision
  nodes are labelled. These labels tell you what decision is made there.

- Chance nodes Represented by a circle, a chance node identifies where
  in the decision tree uncertainty exists. The children of the chance
  node represent different outcomes and each chance node has its own
  probability distribution.

- Utility nodes Utility nodes are diamond shapes and are the leafs in
  the tree. They represent the outcomes and have a score assigned to
  them.

Decision trees can be used to map every possible scenario in a decision
problem. Here are some of its distinguishing features. [16, 30]

Allow the possibility of additional scenarios

If it turns out more choices can be made during a decision problem this can
be updated in the model by adding additional levels to the tree.
This could be useful if hospital policy were to add an additional decision to
be made during certain treatments.

Easily interpreted

Due to its structure it is not difficult to understand a decision tree. This
makes it easier to convince people to use one.

Different outcomes are more easily compared to each other

It is relatively easy to find the value nodes with the worst and best outcomes.
Also, it is possible to find the path with the best expected outcomes.

Highly detailed

Because decision trees map every decision possible they represent the deci-
sion problem with great detail and allow for very thorough analyses.

3.2.4 Known disadvantage of decision trees

Unfortunately decision trees also have some disadvantages which in certain situations make other network models preferable [16, 30]

Multiple output variables are not allowed

In order to use a decision tree all variables used within must be reduced to a single variable. Usually utility.

High complexity

When the tree expands the number of calculations needed to perform certain operations increase exponentially. While this is not a problem for small trees with large trees with a lot of unknown variables in the form of chance nodes the computational costs could be very high. In case of a complex model such as a medical decision problem this is a possible occurrence. Though one might attempt to reduce this by pruning the tree a little this is usually not enough to counteract the disadvantage.

3.3 Influence diagrams

3.3.1 Formal definition of influence diagrams

Formally an influence diagram $I = (V(I), A(I))$, which is a directed acyclic graph in which V(I) is a set of vertices and $A(I)$ is a set of directed edges.

$V(I) = C(I) \cup D(I) \cup U(I)$ with $C(I), D(I), U(I)$ mutually disjoint where $C(I)$ is a set of chance nodes of which each has its own probability distribution with a chance assigned to each eventuality, $D(I)$ is a set of decision nodes and $U(I)$ is a set of utility nodes.

Influence diagrams are also probabilistic graphical models and therefore also are acyclic and directed. Unlike Bayesian networks however they do not solely consist of chance nodes but like the decision trees they may also contain decision nodes and utility nodes.

3.3.2 Explanation of a simple influence diagram

P(Rainy|Rain) = 0.8
P(Sunny|Rain) = 0.2
P(Rainy|Dry) = 0.2
P(Sunny|Dry) = 0.8

U|Rain, Indoor = 6
U|Rain, Outdoor = 1
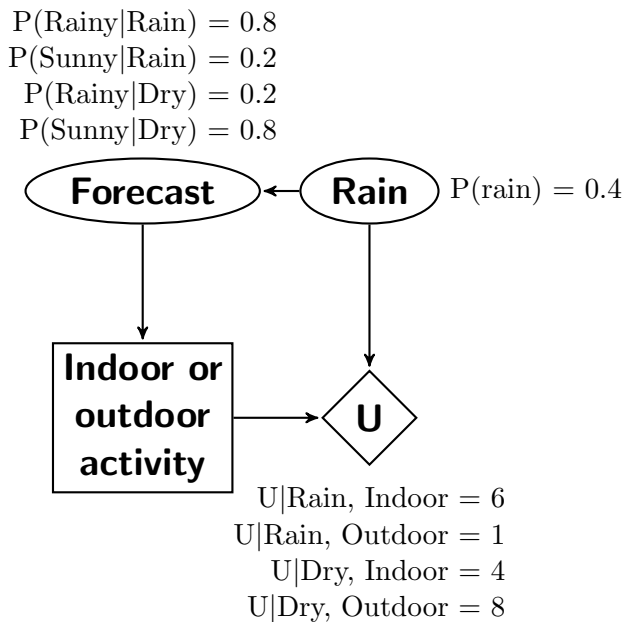U|Dry, Indoor = 4
U|Dry, Outdoor = 8

Figure 3.3: An influence diagram

As we can see there is a chance node labelled Rain which shows that the chance that the actual chance that it will rain during the activity is

0.4. According to the chance node labelled Forecast the weather forecast is influenced by the weather during the activity and that the forecast has a chance of 0.8 to predict the weather correctly. The decision node shows us we have the choice of doing an indoor or an outdoor activity and the value node shows us the utility for each outcome. Should the forecast say that it will be sunny we know that the likelihood of the scenario's involving rain is 0.2 and those involving dry weather 0.8, therefore we can calculate the expected utilities for the indoor and outdoor activities. For the indoor activity it would be 6 * 0.2 + 4 * 0.8 = 4.4. For the outdoor activity it is 1 * 0.2 + 8 * 0.8 = 6.6. The model suggests that if the forecasts says it will be sunny selecting the outdoor activity has the highest expected utility value and should be chosen.

3.3.3 Features of influence diagrams

Influence diagrams are directed acyclic graphs. Like decision trees they have chance, decision and utility nodes. Features that make them suitable to be used as decision support systems for clinical treatment selection are as follows [15, 16].

Highly modular

A large influence diagram could be divided into smaller sub diagrams which allows focusing on one aspect of a decision problem. Also, due to its modularity, influence diagrams can be maintained and updated more easily.

Compact graphical depictions of a decision problem

Influence diagrams allow for networks with considerably less nodes than decision trees. This means it is easier to display them. Also, they exhibit less growth as the size of the system increases compared to the other models.

3.3.4 Known disadvantages of influence diagrams

Accurately constructing a diagram is difficult

Compared to the relatively simple decision tree and the somewhat simpler Bayesian network it may be difficult to accurately model all the causal relations with accurate probability distributions and all the possible decisions into an influence diagram. Especially with complex decision problems. [15].

Solution algorithms have a high complexity

Though graphically an influence diagram appears small enough, obtaining the decision with the best expected outcome requires a lot of calculations.

3.4 How is a model made?

3.4.1 The basic steps of model building

Though there are many approaches to creating models there are a few rough steps that most approaches have in common [10]. These steps are the following.
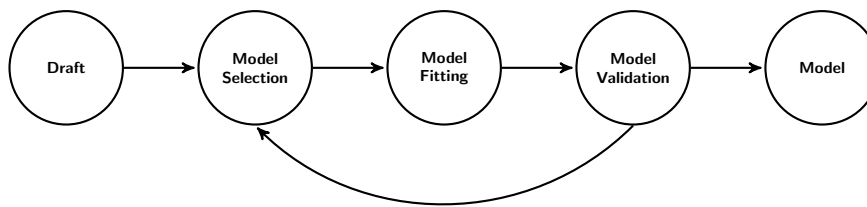

Figure 3.4: The steps of model building

1. **Draft** First it has to be established that there is a need for a model. The assumptions regarding it and any constraints have to be listed and data on which it will be based has to be obtained.

2. **Model selection** In this step the available data such as patient information, possible diagnosis results and knowledge regarding the medical processes involved are used to select the form of model that would allow the data to be fitted in there.

3. **Model fitting** The data is fitted into the model. For example the different parameters get their appropriate probability distributions assigned to them if known. Unknown parameters are being estimated as best as possible or, through machine learning, could be trained.

4. **Model validation** The model is being evaluated. If during model validation some problems are detected then the model can be improved by going back to step one. Should no problems be detected then model is ready for the next phase.

3.5   Metrics used when validating

Regardless of the data the model is being validated with, there will be a variety of metrics available.

Accuracy

Accuracy is how closely the answer a network model provides approaches reality. This could be calculated by comparing the decisions a network system makes to the actual decisions made. In case of a classifying network model it the classifications the model makes would be compared to the actual classes. The number of correct decisions or classifications would be divided by the total number of measurements to give us the accuracy. In short, it gives us the fraction of the correct classifications or decisions:

$$(tp + tn)/(tp + fp + fn + tn)$$

Precision

In case of classification precision is the number of instances correctly classified as a certain class divided by the total number of instances classified as that class:

$$tp/(tp + fp)$$

Recall

In classification recall is the number of instances belonging to a class that are classified as such divided by the total number of instances belonging to that class. This is also referred to as sensitivity

$$tp/(tp + fn)$$

ROC

ROC stands for receiver operating characteristic or ROC curve. The ROC curve is a graphical plot where the true positive rate is plotted against the false positive rate at several threshold settings, resulting in a curve. The steeper the curve, the better the model is. Also, there is the ROC space.

Here the true positive rate and the false positive rate are also plotted against each other but the results will be expressed as a point in that plot rather than a curve.

Brier score

The Brier score is a score function that can be used when the predictions assign probabilities to mutually exclusive outcomes. For example, a patient will survive, or a patient will not survive.
The Brier score is most commonly formulated as follows""

$$BS = \frac{1}{N} \sum_{t=1}^{N} (f_t - o_t)^2$$

As the formula shows the Brier score is calculated by summing up the squared differences between predicted result and observed result and then dividing that by the number of observations. A lower Brier score means that the model has yielded better predictions.

Log-likelihood ratio

The log-likelihood ratio is used to compare the goodness of fit of two models. This is done by using a set of observations that we have. We know those to be true so we can calculate how likely these observations are to occur given the two models. The probability for these observations to occur in each model are compared by dividing the likelihood of the model to be evaluated by the likelihood of the model it is compared to, the null model. We apply log to the resulting ratio and the higher the result the better the goodness of fit of the model is compared to the null model.

Chapter 4

Clinical Treatment Selection

Clinical treatment selection is essentially a decision-making problem [6, 25, 29]. In order to select the treatment selection that leads to what a physician believes to be the most desirable result he must have gone through certain processes.

First a patient has to present him or herself with some signs that something might be wrong with the patient. This can happen in several ways. It could be that during a routine check up the examining physician was unable to confirm the absence of pathological processes and decided that a more thorough examination is required. It could also be that a patient visits his or her physician with a complaint regarding his or her health. This is where the diagnostic procedure begins.

The diagnostic procedure

At this point there is too much uncertainty to make a proper decision. Therefore the physician will need to reduce the uncertainty. Usually this procedure begins with the physician methodically gathers information in the form of patient data. This is called an anamnesis. The anamnesis plus the signs the patient already exhibits usually give the physician some ideas of what might be wrong with the patient. Physicians tend to work hypothesis driven, therefore he is going to methodically try to eliminate as many of the possibilities as he can by asking further questions and performing more diagnostics. Ideally the physician would have eliminated all but one of the possibilities in this manner in which case only one disease needs to be treated. It is possible however that there are still several possibilities. Some diagnostic actions the physician may resort to are

- Wait and observe By allowing the disease to take its course it may become more evident which one the patient is suffering from. The advantages of this choice are that no invasive diagnostic actions have to be performed and more certainty is gained. The disadvantage is that the disease is allowed to progress further.

- Use more invasive diagnostic procedures More information might be gained by performing diagnostic procedures which are more invasive. Such procedures include explorative surgery or a highly detailed CAT scan. Depending on how invasive the diagnostic procedure is there will be some damage caused to the patient which is a clear disadvantage. An advantage is that it might allow for more certainty at an earlier stage which might prevent other damage.

Treatment selection

At a certain point a treatment will be selected. This could be because further diagnostic procedures will not be beneficial or simply because there is no more time to do anything else but treat the patient. Usually a prognosis is made based on the current status of the patient. The physician will use his medical knowledge to make educated guesses what the influence of certain treatment options may have on the prognosis of the patient. By comparing the different expected outcomes of each of the treatment options the physician is going to choose the treatment which will lead to the best expected prognosis.

Prognostic process

Once the patient has undergone the selected treatment it is customary to monitor the patient to see if the treatment was successful. With more data available now regarding how the treatment went a new, more accurate prognosis can be made. If this prognosis differs too greatly from what was expected during treatment selection in an undesirable way the patient may have to undergo additional treatment.

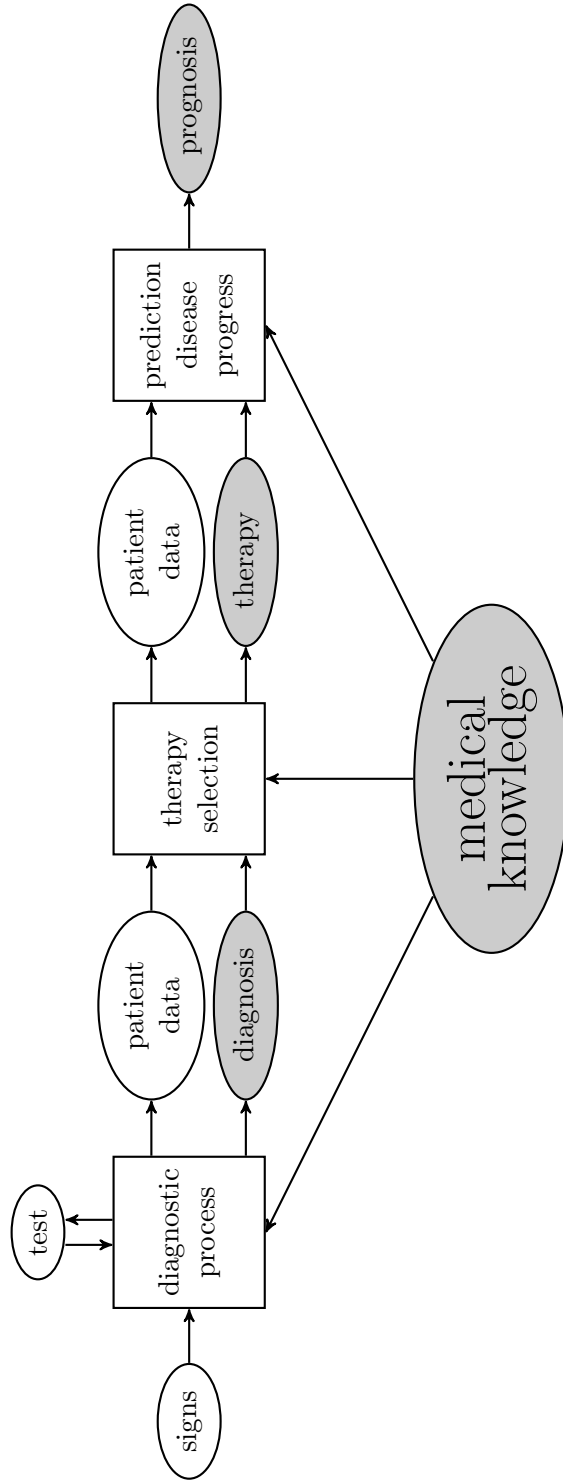Figure 4.1 shows schematically how clinical treatment selection works.

Figure 4.1: Overview of the medical procedures

Ways in which network models might support a physician

How could network models provide support to a physician? First we need to briefly establish that for each of the stages as shown in the above illustration there is a possible model that could support a physician [22].

- Diagnostic models can help a physician interpret the data collected during diagnostic actions [21]. Basically it can be calculated how much uncertainty there currently is and how much more certainty is gained by performing an additional diagnostic action, therefore having a more accurate estimation of when it is time to select a treatment.

- A decision model can help a physician decide if enough certainty has been gained to commence treatment or that more certainty is necessary and another diagnostic action must be performed to collect more data [21]. It can also help choose which treatment to select by calculating which decision with the current evidence will yield the highest expected utility.

- A prognostic model can give the clinician an overview of the possible outcomes and their likelihoods using the available patient data and diagnostic information as evidence [1–3].

Now if we combine the three models like in the illustration below we can create a model that can function as a decision support system.

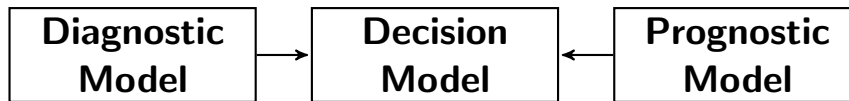| **Diagnostic Model** | → | **Decision Model** | ← | **Prognostic Model** |

Figure 4.2: Flow of information in a network model designed for clinical decision making

Basically the resulting network model would use the information gained from the diagnostic model to see which treatment options are available. The information from the prognostic model would yield information about the likely result of each possible decision. The combined model can therefore pick the decision with the best expected outcome.

This network model could basically be plugged into the figure about the medical treatment process to illustrate its supportive function.
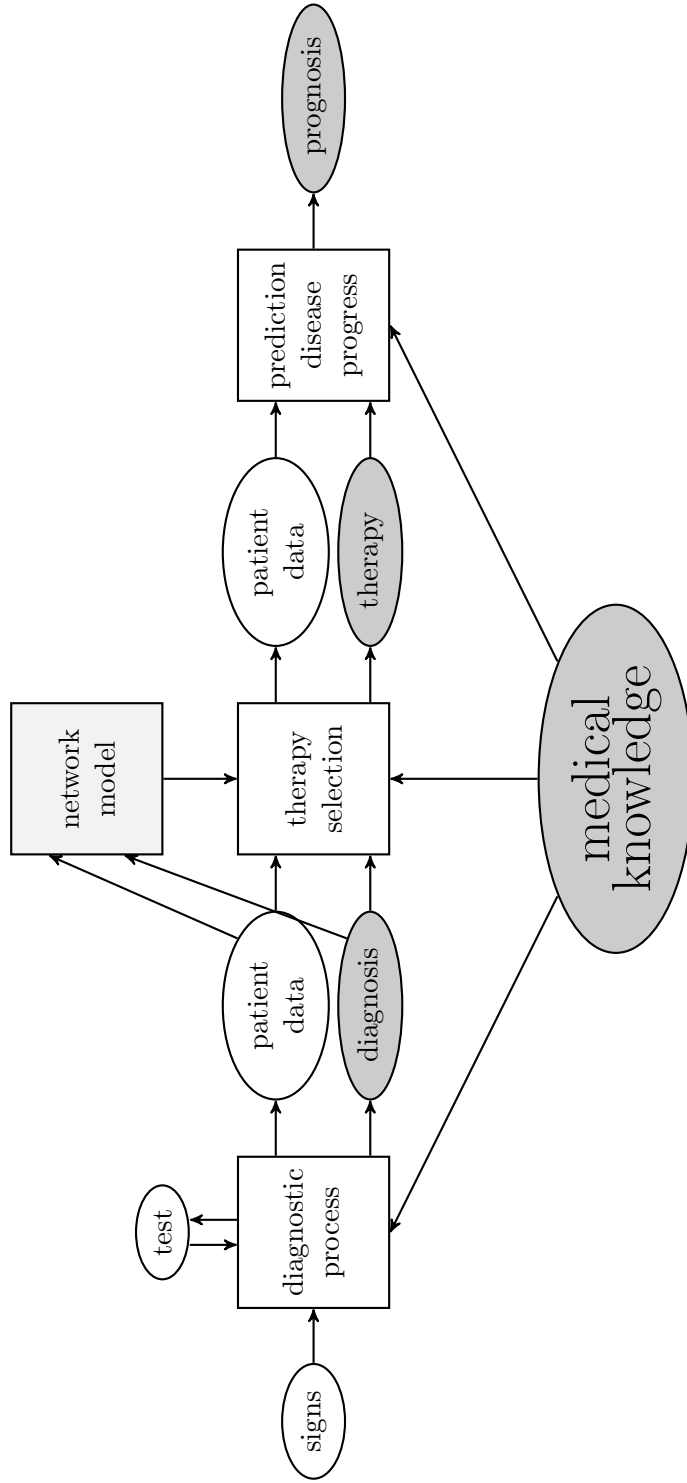
Figure 4.3: Overview of the medical procedures with a network model plugged in

Chapter 5

Evaluation Methods

5.1 Why Evaluate?

There are several reasons why we would evaluate network systems used for clinical decision making [12]

1. To stimulate physicians to use these network models. In order to promote the use of these models we must be able to show the advantages of using them. Should a model get a positive evaluation this is more easily done.

2. To further advance science. These models are constantly being developed. Thoroughly evaluating them may yield new insights which would lead to better results.

3. To aid developers of network models. Without a way to properly evaluate their work, developers will not know if they did a good job or not. Proper evaluation will help developers become more effective at their work.

4. To justify the use of the network model. If it hasn't been evaluated then it is unknown if a network model is worth the cost of implementing it because it is unknown how well it does compared to other resources.

5. To reduce the risk of liability. If a patient has received a less than optimal treatment and it turns out that the network model consulted had not been properly evaluated prior to being implemented the patient may decide to sue.

5.2 Difficulties when evaluating a model

Evaluating a network model meant to assist in clinical treatment selection will not likely be a very straightforward task. A medical decision is usually the result of a combination of objective and subjective variables [6, 25, 29]. The objective variables are objectively measurable with a very low degree of inaccuracy such as weight, height, age, size, family history.

There are however more subjective variables which are also taken into account such as discomfort, quality of life. Those are very difficult to measure in a uniform manner that holds true for multiple patients because obviously every patient is unique and therefore would have assigned different numeric values to certain degrees and forms of discomfort and would different factors would influence their opinion about their quality of life differently.

Also there are those variables that although seemingly quantifiable are still subjective due to the measured value being dependent on subjective things or somehow influenced by factors outside of the scope of the model. Cost of the treatment expressed in Euros for example is one of these because although the cost in Euros is objectively measured the difference in cost cannot always be rationally explained. The same treatment could have a different cost depending on the facility the treatment will take place. Also, it would seem that when ones health is concerned one tends not to assign much weight to the cost of the treatment. This has caused a lot of things medical to be of a higher price than they usually would be and others not which makes cost in Euros a less reliable measure.

Another such variable would be life expectancy. Though it is measurable, only when the patient has died will one know what his or her actual life expectancy was. Not only is that a factor, but also the fact that there are a great many variables outside of the scope of the model that influence this variable.

Because people are involved it also remains to be seen how closely they follow the network model. It might be that the physician disagrees with the model to such an extend that he decides to select a different treatment. Of course it would still be possible to check that models prediction regarding the actual treatment chosen [5, 7].

People also tend to behave differently when they know the are being evaluated or are being part of an evaluation [12]. This change to less natural behaviour may influence the observations made during the evaluation and it is hard to determine in what way and to which degree precisely.

5.3   What is there to evaluate in a network model?

There are quite a number of things one can examine in a network model. First I shall go through some of the most important metrics and definitions.

5.3.1   Validation

Validity is how closely a model represents the real system [2, 3]. Validation therefore involves checking how accurately the model represents this real system. There are several kinds of validation. In the next few sections I will explain ways in which a network model can be statistically validated.

Internal Validation

Internal Validation is the process of evaluating whether the model is valid when only previously acquired data is used and no new data is introduced. A medical research facility may use the data they collected on patients in a single year and base their model only on that information. Internal validation checks the model using only that data.

One could check the model by using a sample of the data to see if the model yields a correct prediction or decision. If that data has been used by the model to learn however it can be assumed to give the correct answer. It is impossible to say anything about the predictive value of the model using this approach since the model just looks up the cases in the knowledge it has and is guaranteed to have an exact match.

To properly apply internal validation on a network model one would want to use cross-validation. The data is divided into training data and test data. The training data is used to train the model and the test data is used to test the predictive value of the model. Depending on how exhaustive one wants to test the model and the data one can adjust the ratio between training and test data and the number of fold in which the data is divided. Cross validation is a method to prevent overfitting, thus making sure that the perceived predictive value will not turn out to be very different from the actual predictive value.

External Validation

External validation involves testing the model against data other than the data initially used to build the model. One could test it on patient data obtained by another medical research centre for example. If the model

performs well on external data as well its external validity is determined. An externally valid model is considered applicable on a more general population.

Temporal Validation

Temporal validation is applied when the same research centre evaluates the performance of the model over time. The population will be the same for the most part with a few different patients possibly. Technically it is somewhat in between internal and external validation in the sense that most of the predictions or decisions will be based on cases exactly the same as in the training data but that there will also be some new data to test it against. Not as much as when externally validated though.

The difference between validation and evaluation

Statistical validation is purely concerned with how well a model fits its data. Clinical validation is about whether the model is relevant and usable. Evaluation is of a broader scope and could also involve analysing how reliable the test was and what the long term effect of the model might be on healthcare. Evaluation may also have a broader scope in regard to purpose.

5.4   When is a model evaluated?

During its lifespan a model is going to get evaluated a number of times. It could very well be that it is constantly being evaluated. I shall give a list in chronological order of these evaluation moments and which methods of evaluation are of most use.
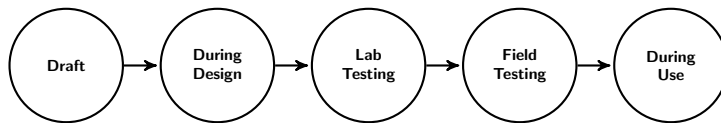

Figure 5.1: Lifespan of a model

Draft

Already during it's draft stage a few things can be taken into consideration. If the model that is being considered to make does not have sufficient data to fit a model to then the nest stage may have to be postponed until enough

data has been gathered. If it doesn't seem like the model would be useful then aborting its design may need to be considered.

Design

While the model is in this early of a stage it can be checked if the assumptions made while selecting the model and fitting the data make any sense. If the model represents the domain knowledge relevant to its function properly. If the model is going to be useful should it be implemented as is. And of course if the model is going to do what it is being designed for. In case of a purely diagnostic model it can be checked if the model drew the correct conclusion. In case of a binary scale such as if a patient tested positive or negative for a disease we can use accuracy, precision or recall to see how it performs.

In case of a model that functions as a classifier we can evaluate how well it classifies by measuring the classification error. Usually this is done using precision and recall.

For a continuous scale model area we could use residual or R^2. If the differences between predicted and observed value seem random and the difference isn't so much it can no longer be explained by random measurement error then the model was fitted sufficiently. If there seems to be too much of a recurring difference then the model may have to be refitted.

Evaluating diagnostic models seems the easiest since a diagnosis is either false or correct. A complicating factor for a prognostic model is that in order to validate it the predictions have to be checked. This is only possible if the data regarding these predictions is available. For example, if the model predicts whether the patient will survive the procedure one will have the answer relatively quickly. If the model tries to predict survival after five years it may take five years to validate the model.

Decision support models, since they combine aspects of the diagnostic and prognostic models have the same problem when being evaluated. If one wishes to validate a model quickly it would therefore be recommended to do a retrospective analysis so that all the required data has been collected in advance.

At this stage pretty much every form of validation will involve internal validation because the model is being checked against the data it was based on.

Once the model has been validated internally it should be checked if it is applicable on new data. This can be done by applying the same methods of validation on the model but with different data. For temporal validation

37

it would use data from the same research centre but accumulated during a different time interval. Data collected one year later for example. For external validation completely new data of another population would be used. A slightly decreased performance is to be expected. If the model performs badly however this is likely to be a sign of overfitting in which case the model works nicely for the dataset it was trained on but is not quite applicable for the general population. It may require more data or the way it trains on its data may have to be reviewed.

Lab test

When the model has also been validated externally and as much as possible with retrospective studies it is time to use it. It is possible to do a lab test which would allow for more control over the parameters due to the controlled environment provided by a laboratory. Once again the model will be validated by measuring the differences between its predicted results and the observed results. The same measures would be used as in the previous steps except that in the lab one can also test the usability of the model and perhaps how satisfactory its usage is for the clinicians and patients. Factors to keep in mind are that the people involved in the experiment know they are being observed and thus may act differently than they otherwise would. Also, for the lab test to be useful it has to be an approximately realistic setting.

Field test

Lastly the model can be field tested. This would involve using the model as it was intended from the beginning. As a decision support system used by clinicians on actual patients. Statistically the model can be evaluated in the same manner. There are however other aspects of the model and its implementation that can be evaluated now or that have to be factored in. Once again the Hawthorne effect will play a role. Also, the results will be influenced by the compliance of the clinicians and patients. Though everyone is technically free to make their own decisions for the sake of evaluating the model it must be documented when it is decided not to follow the model.

Implementation

Of course it is recommended that the model is periodically evaluated once it has been accepted and implemented. This should be part of its maintenance routine. Evaluating it periodically allows for measurement of the

improvement of healthcare and can yield new insights concerning the decision problem which might further improve the model.

Chapter 6

Conclusions

Thanks to the abundance of literature available about related subjects such as network models used for diagnostic models, prognostic models, statistical validation of models, clinical validation, medical decision analysis, probability theory and evaluation it was possible to find methods of evaluation applicable to network models used in clinical treatment selection.

Evaluating such a model would ideally occur several times during its design. It would get statistically validated so that it fits well on the data used to create it. Then one would proceed to see how it handles external data. Once the model works sufficiently well on external data it can be tested in a lab environment. If it performance is satisfactory and it is also determined that it is relevant and not too difficult to use it can be subjected to a field test. If under these less controlled circumstances it still performs well it should be ready for implementation. Of course, once it is implemented it can still be periodically evaluated as part of its maintenance. Also, it can now be measured if it is of influence on the performance of the facility it is implemented in and on a larger scope, healthcare in general could be measured as well.

It may take some time to evaluate some of the models and it is also dependent on the compliance of physicians and patients, but by using the proposed method to evaluate network models used in clinical treatment selection we could see some increased usage of these models. Of course additional research regarding the actual implementation of these models and how they perform as well as regarding how willing physicians are to use these models would be desirable.

Bibliography

[1] Ameen Abu Hanna, Peter JF Lucas, et al. Prognostic models in medicine-ai and statistical approaches. *Methods of Information in Medicine*, 40:1–5, 2001.

[2] Douglas G Altman and Patrick Royston. What do we mean by validating a prognostic model? *Statistics in medicine*, 19(4):453–473, 2000.

[3] Douglas G Altman, Yvonne Vergouwe, Patrick Royston, and Karel GM Moons. Prognosis and prognostic research: validating a prognostic model. *Bmj*, 338:b605, 2009.

[4] Peter Antal, Geert Fannes, Dirk Timmerman, Yves Moreau, and Bart De Moor. Using literature and data to learn bayesian networks as clinical models of ovarian tumors. *Artificial Intelligence in medicine*, 30(3):257–281, 2004.

[5] Sawsan Baddar, Elizabeth A Worthing, Omar A Al-Rawas, Youssef Osman, and Bazdawi M Al-Riyami. Compliance of physicians with documentation of an asthma management protocol. *Respiratory care*, 51(12):1432–1440, 2006.

[6] Patrick Bossuyt. *Klinische besliskunde*. Amsterdam University Press, 1994.

[7] Dror Cantrell, Oded Shamriz, Matan J Cohen, Zvi Stern, Colin Block, and Mayer Brezis. Hand hygiene compliance by physicians: Marked heterogeneity due to local culture? *American journal of infection control*, 37(4):301–305, 2009.

[8] Brian J Cohen. Is expected utility theory normative for medical decision making? *Medical Decision Making*, 16(1):1–6, 1996.

[9] Robert G Cowell. *Probabilistic networks and expert systems: Exact computational methods for Bayesian networks.* Springer Science & Business Media, 2006.

[10] Carrol Croarkin and Paul Tobias. *NIST/SEMATECH e-Handbook of Statistical Methods.* NIST, 2003.

[11] M Julia Flores, Ann E Nicholson, Andrew Brunskill, Kevin B Korb, and Steven Mascaro. Incorporating expert knowledge when learning bayesian network structure: a medical case study. *Artificial intelligence in medicine*, 53(3):181–204, 2011.

[12] Charles P Friedman and Jeremy Wyatt. *Evaluation methods in biomedical informatics.* Springer Science & Business Media, 2005.

[13] Sven Ove Hansson. Decision theory – a brief introduction, 1994.

[14] Cecil Huang and Adnan Darwiche. Inference in belief networks: A procedural guide. *International journal of approximate reasoning*, 15(3):225–263, 1996.

[15] James E. Matheson James E. Smith, Samuel Holtzman. Structuring conditional relationships in influence diagrams. *Operations Research*, 41(2):280–297, 1993.

[16] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning.* The MIT Press, 2009.

[17] A.N. Kolmogorov. *Foundations of the Theory of Probability.* Martino Fine Books, 2013.

[18] Kevin B Korb and Ann E Nicholson. *Bayesian artificial intelligence.* CRC press, 2010.

[19] Dimitris Margaritis. *Learning Bayesian network model structure from data.* PhD thesis, US Army, 2003.

[20] David Meltzer. Accounting for future costs in medical cost-effectiveness analysis. *Journal of health economics*, 16(1):33–64, 1997.

[21] RA Miller. Medical diagnostic decision support systems–past, present, and future: a threaded bibliography and brief commentary. *Journal of the American Medical Informatics Association*, 1(1):8, 1994.

[22] Yves Moreau, P Antal, Geert Fannes, Bart De Moor, et al. Probabilistic graphical models for computational biomedicine. *Methods of information in medicine*, 42(2):161–168, 2003.

[23] Sucheta Nadkarni and Prakash P Shenoy. A causal mapping approach to constructing bayesian networks. *Decision support systems*, 38(2):259–281, 2004.

[24] Thomas Dyhre Nielsen and Finn Verner Jensen. *Bayesian networks and decision graphs.* Springer Science & Business Media, 2009.

[25] Harold C Sox. *Medical decision making.* ACP Press, 1988.

[26] Jiang Su, Harry Zhang, Charles X. Ling, and Stan Matwin. Discriminative parameter learning for bayesian networks. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 1016–1023, New York, NY, USA, 2008. ACM.

[27] Marcel Antonius Johannes van Gerven. *Bayesian networks for clinical decision support: a rational approach to dynamic decision-making under uncertainty.* [Sl: sn], 2007.

[28] John Von Neumann and Oskar Morgenstern. *Theory of games and economic behavior.* Princeton university press, 2007.

[29] M.C. Weinstein and H.V. Fineberg. *Clinical Decision Analysis.* Saunders, 1980.

[30] Yongheng Zhao and Yanxia Zhang. Comparison of decision tree methods for finding active objects. *Advances in Space Research*, 41(12):1955–1959, 2008.