Artificial Intelligence and Bayesian Network

- Intelligence, Application, Criteria for judging success
- AI - computations to perceive, reason and act
- Engineering goal - solve real world problem using AI as an armamentarium of ideas about representing knowledge, using knowledge and assembling systems
- Scientific goal - to determine which ideas about representing knowledge, using knowledge and assembling systems explain various sorts of intelligence
- in farming - controlled robots - control pests, prune trees, selectively harvest mixed crops

Artificial Intelligence and Bayesian Network

- in manufacturing - robots - do inspection, maintenance job, dangerous and boring assembly
- in household work - advice on cooking, cleaning, shopping, do laundry
- in medical care - help practitioners with diagnosis, monitor patient's condition, manage treatment
- in school - computers act as superbooks, helping students to understand the topics, provides answer to question
- AI helps in analysis, synthesis, learn from examples, experience or data
- AI more essential

Artificial Intelligence and Bayesian Network

- Airlines - allocate gate to arriving flights, schedule departure, avoid potential traffic jam, catering, passenger service, crew scheduling, aircraft maintenance
- representation in artificial intelligence
- representation is a set of conventions about how to describe a class of things
- description makes use of conventions of a representation to describe some particular thing
- finding appropriate representation is a major part of problem solving

Artificial Intelligence and Bayesian Network

- example, farmer wants move a fox, a goose and grain across river his boat tiny he can take only one of his possessions across any trip
- unattended fox will eat a gooses and unattended goose will eat grain
- farmer must not leave fox alone with goose or goose alone with grain
- question is what he should do?
- English is not a good representation
- need to separate important constraints from irrelevant details
- node for each farmer and his three possessions, two banks of river
- $2^{1+3} = 16$ arrangements
- 10 of which are safe in the sense that nothing is eaten

Artificial Intelligence and Bayesian Network

- six unsafe arrangements place an animal and something the animal likes to eat on one side with the farmer on the other
- draw a link for each allowable boat trip
- for each ordered pair of arrangements there is a connecting link if and only if the two arrangements meet two conditions: first farmer changes sides and second at most one of the farmer's possessions changes sides
- there are 10 safe arrangements and there are 10 x 9 =90 ordered pairs but only 20 of these pairs satisfy the conditions required for links
- node-and-link description is a good description with respect to the problem posed
- good description, developed within the conventions of a good representation, leads to problem solving

Artificial Intelligence and Bayesian Network

- a problem is described using an appropriate representation, the problem is almost solved
- good representation, expose natural constraints, able to express one object or relation influences another
- suppress irrelevant details,
- representation should be evaluated based on
  - transparent - understand what is being said,
  - complete - say all that needs to be said,
  - concise - say what you need to say efficiently,
  - fast - store and retrieve information rapidly,
  - commutable - create using existing procedure

Artificial Intelligence and Bayesian Network

- representation - consists of four parts
- lexical part determines which symbols are allowed in the representation's vocabulary
- structural part - describes constraints on how the symbols can be arranged
- procedural part - specifies access procedures that enable you to create descriptions, to modify them and to answer questions using them
- semantic part - that establishes a way of associating meaning with descriptions
- lexical - nodes and links
- structural - links connect node pairs
- procedural - it is in brain for farmer example

Artificial Intelligence and Bayesian Network

- semantic - establishes nodes correspond to arrangements of the farmer and his possessions and links correspond to river traversals
- semantic nets convey meaning
- semantic nets consist of nodes, denoting objects, links, denoting relations between objects and link labels that denote particular relations
- from semantic perspective the meaning of nodes and links depends on the applications
- semantic nets - examples - semantic tree - search tree, decision tree, goal tree, game tree
- it may state space or frame system - value propagation net, constraint net

Artificial Intelligence and Bayesian Network

- problem solving and understanding knowledge
- knowledge concerns the description of concrete or abstract objects
- knowledge may fit within the semantic net frame work or embedded in procedures
- how much knowledge required?
- rule base system - inference net
- example, rule chaining led to conclusion a particular animal is giraffe and another led to verification of hypothesis that a particular animal is a cheetah
- rule based systems viewed as models
- rules are called productions

Bayesian Artificial Intelligence

- incorporating Bayesian inferential method in development of AI based system
- learning from observation and experiment
- knowledge engineering with Bayesian networks
- Bayesian methods: classification, curve fitting, time series modeling
- AI - intelligence developed by humans
- Alan Turing's famous test — the ability to fool ordinary (unfoolish) humans about whether the other end of a dialogue is being carried on by a human or by a computer
- automation, transform our technology and economy

Bayesian Artificial Intelligence

- AI system -
  - deals with uncertainty,
  - deals with incomplete evidence leading to beliefs
  - that fall short of knowledge, with fallible conclusions and
  - the need to recover from error,
- called non-monotonic reasoning
- AI will need to be able to reason probabilistically, called Bayesian reasoning

Bayesian Artificial Intelligence

- three distinct forms of uncertainty which an intelligent system
- Ignorance - the limits of our knowledge lead us to be uncertain about many things
- Physical randomness or indeterminism - toss a coin,
- there will remain an inescapable degree of uncertainty about outcome head or tail
- Vagueness - many of the predicates, employed appear to be vague
- it is often unclear whether to classify a dog as a spaniel or not, a human as brave or not, a thought as knowledge or opinion

Bayesian Artificial Intelligence

- Bayesianism is the philosophy that asserts that in order to understand human opinion
- as it ought to be, constrained by ignorance and uncertainty,
- the probability calculus is the single most important tool for representing appropriate strengths of belief
- Bayesian computational tools for reasoning with and about strengths of belief as probabilities
- Bayesian view of physical randomness
- knowledge representation and reasoning within AI

Bayesian Artificial Intelligence

- probability calculations are hard — in fact, NP hard in the number of variables — called “naive Bayes”
- certainty factors - degree of belief and disbelief
- rule based system - allowing rules to be modular, combined impact of diverse evidence being a compositional function of their separate impacts
- Bayesian networks provide a natural representation of probabilities
- understanding and representing uncertainty with probabilities
- probability calculus allows us to represent the independencies which other systems require, but also allows us to represent any dependencies which we may need

Bayesian Artificial Intelligence

- Bayes' Theorem - theorem of the probability calculus
- the probability of a hypothesis h conditioned upon some evidence e is equal to its likelihood $P(e|h)$ times its probability
- prior to any evidence $P(h)$, normalized by dividing by $P(e)$ (so that the conditional probabilities of all hypotheses sum to 1)
- adjusting our beliefs in our hypotheses given new evidence is called conditionalization
- after applying Bayes's theorem to obtain $P(h|e)$ adopt that as posterior degree of belief in h or $Bel(h) = P(h|e)$
- belief updating via probabilities conditional upon the available evidence
- it identifies posterior probability - the probability function after incorporating the evidence

Bayesian Artificial Intelligence

- there are certainly situations where conditionalization very clearly does not work
- the two most basic such situations simply violate what are frequently explicitly stated as assumptions of conditionalization:
  1. there must exist joint priors over the hypothesis and evidence spaces - without a joint prior, Bayes' theorem cannot be used, so conditionalization is a non-starter
  2. the evidence conditioned upon, e is all and only the evidence learned - called the total evidence condition, - significant restriction, since in many settings it cannot be guaranteed

Bayesian reasoning examples

- Breast Cancer
- suppose the women attending a particular clinic show a long-term chance of 1 in 100 of having breast cancer
- suppose also that the initial screening test used at the clinic has a false positive rate of 0.2 (that is, 20% of women without cancer will test positive for cancer) and that
- it has a false negative rate of 0.1 (that is, 10% of women with cancer will test negative)
- the laws of probability dictate from this last fact that
- the probability of a positive test given cancer is 90%
- suppose that there is such a woman who has just tested positive

Bayesian reasoning examples

- what is the probability that you have cancer?

$$P(Cancer|Pos) = \frac{P(Pos|Cancer)P(Cancer)}{P(Pos)}$$

- $P(Pos|Cancer)$ = the probability of Pos given Cancer — which is the likelihood 0.9
- $P(Pos) = P(Pos|Cancer)P(Cancer) + P(Pos|\neg Cancer)P(\neg Cancer) = 0.9 \times 0.01 + 0.2 \times 0.99 = 0.009 + 0.198$
- $P(Cancer|Pos) = \frac{0.9 \times 0.01}{0.009+0.198} \approx 0.043$
- the discrepancy between 4% and 80 or 90%

Bayesian Artificial Intelligence

- interested in knowing:
- how humans perform the many interesting and difficult cognitive tasks encompassed by AI
- such as, natural language understanding and generation, planning, learning, decision making
- but also interested in knowing how they might be performed otherwise, and in knowing how they might be performed optimally
- by building artifacts which model our best understanding of how humans do these things
- called descriptive artificial intelligence and also
- building artifacts which model our best understanding of what is optimal in these activities
- called normative artificial intelligence

Bayesian Artificial Intelligence

- understanding of the nature of intelligence and also produce some very useful tools for science, government and industry medical, legal,
- most other varieties of human reasoning either consider the relevant probabilistic factors and accommodate them or run the risk of introducing egregious and damaging errors
- the goal of a Bayesian artificial intelligence is to produce a thinking agent
- which does as well or better than humans in such tasks, which can adapt to stochastic and changing environments
- recognize its own limited knowledge and cope sensibly with these varied sources of uncertainty
- how can it be achieved?
- the first step is to develop algorithms for doing Bayesian conditionalization properly

Bayesian Artificial Intelligence

- difficulties in developing applications, difficulties with eliciting knowledge from experts, and integrating and validating the results
- issue is that there is no clear methodology for developing, testing and deploying Bayesian network technology in industry and government
- there is no recognized discipline of "software engineering" for Bayesian networks
- reasons for artificial intelligence to use probabilistic reasoning
- dealing with probabilities in AI, namely Bayesian networks

Bayesian Artificial Intelligence

- Bayesian networks (BNs) are graphical models for reasoning under uncertainty
- the nodes represent variables (discrete or continuous) and arcs represent direct connections between them
- direct connections are often causal connections
- BNs model the quantitative strength of the connections between variables
- allowing probabilistic beliefs about them to be updated automatically as new information becomes available

Bayesian Artificial Intelligence

- how Bayesian networks are put together (the syntax) and
- how to interpret the information encoded in a network (the semantics)
- how to model a problem with a Bayesian network and the types of reasoning that can be performed
- Bayesian network is a graphical structure that allows us to represent and reason about an uncertain domain
- the nodes in a Bayesian network represent a set of random variables from the domain
- a set of directed arcs (or links) connects pairs of nodes, representing the direct dependencies between variables

Bayesian Artificial Intelligence

- assuming discrete variables, the strength of the relationship between variables is quantified by conditional probability distributions associated with each node
- only constraint on the arcs allowed in a BN is that there must not be any directed cycles
- such networks are called directed acyclic graphs, or simply dags
- there are a number of steps that a knowledge engineer (practitioner applying AI technology) must undertake when building a Bayesian network

Bayesian Artificial Intelligence

- Lung cancer: a patient has been suffering from shortness of breath (called dyspnoea) and visits the doctor, worried that he has lung cancer
- the doctor knows that other diseases, such as tuberculosis and bronchitis, are possible causes, as well as lung cancer
- doctor also knows that other relevant information includes whether or not the patient is a smoker (increasing the chances of cancer and bronchitis) and
- what sort of air pollution he has been exposed to
- a positive X-ray would indicate either TB or lung cancer

Bayesian Artificial Intelligence

- nodes and values
- knowledge engineer must identify the variables of interest
- this involves answering the question: what are the nodes to represent and what values can they take?
- nodes that take discrete values
- the values should be both mutually exclusive and exhaustive, which means that the variable must take on exactly one of these values at a time
- Boolean nodes, which represent propositions, taking the binary values true (T) and false (F)

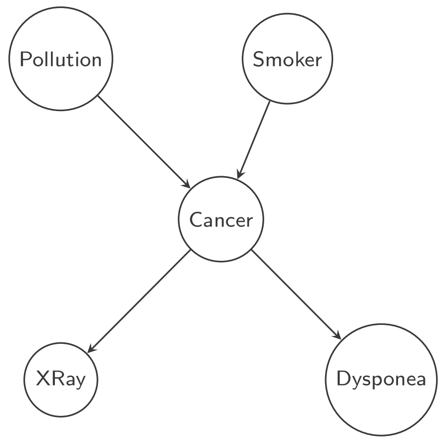
Bayesian Artificial Intelligence

- in a medical diagnosis domain, the node Cancer would represent the proposition that a patient has cancer
- Ordered values - for example, a node Pollution might represent a patient's pollution exposure and take the values $\{low, medium, high\}$.
- Integral values - for example, a node called Age might represent a patient's age and have possible values from 1 to 120
- other diseases, such as TB or bronchitis heavy or a light smoker
- some exposure to pollution for example,

Bayesian Artificial Intelligence

- the structure, or topology, of the network should capture qualitative relationships between variables
- two nodes should be connected directly if one affects or causes the other, with the arc indicating the direction of the effect
- what factors affect a patient's chance of having cancer? if the answer is "Pollution and smoking," then we should add arcs from Pollution and Smoker to Cancer
- similarly, having cancer will affect the patient's breathing and the chances of having a positive X-ray result
- so add arcs from Cancer to Dyspnoea

Bayesian Artificial Intelligence



Bayesian Artificial Intelligence

- a node is a parent of a child, if there is an arc from the former to the latter
- if there is a directed chain of nodes, one node is an ancestor of another if it appears earlier in the chain,
- whereas a node is a descendant of another node if it comes later in the chain
- in example, the Cancer node has two parents, Pollution and Smoker, while Smoker is an ancestor of both X-ray and Dyspnoea
- similarly, X-ray is a child of Cancer and descendant of Smoker and Pollution
- the set of parent nodes of a node X is given by Parents(X)

Bayesian Artificial Intelligence

- Markov blanket of a node, which consists of the node's parents, its children, and its children's parents
- any node without parents is called a root node
- while any node without children is called a leaf node
- any other node (non-leaf and non-root) is called an intermediate node
- root nodes represent original causes, while leaf nodes represent final effects
- in cancer example, the causes Pollution and Smoker are root nodes, while the effects X-ray and Dyspnoea are leaf nodes
- once the topology of the BN is specified, the next step is to quantify the relationships between connected nodes
- this is done by specifying a conditional probability distribution for each node

Bayesian Artificial Intelligence

- for each node, need to look at all the possible combinations of values of those parent nodes
- each such combination is called an instantiation of the parent set
- for each distinct instantiation of parent node values, need to specify the probability that the child will take each of its values
- consider the Cancer node in the figure
- its parents are Pollution and Smoking and take the possible joint values $\{< H, T >, < H, F >, < L, T >, < L, F >\}$
- the conditional probability table specifies in order the probability of cancer for each of these cases to be $< 0.05, 0.02, 0.03, 0.001 >$

Bayesian Artificial Intelligence

- these are probabilities, and must sum to one over all possible states of the Cancer variable,
- the probability of no cancer is already implicitly given as one minus the above probabilities in each case;
- i.e., the probability of no cancer in the four possible parent instantiations is $< 0.95, 0.98, 0.97, 0.999 >$
- the prior for a patient being a smoker is given as 0.3, indicating that 30% of the population that the doctor sees are smokers,
- while 90% of the population are exposed to only low levels of pollution

Bayesian Artificial Intelligence

Node name	Type	Values
Pollution	Binary	$\{low, high\}$
Smoker	Boolean	$\{T, F\}$
Cancer	Boolean	$\{T, F\}$
Dyspnoea	Boolean	$\{T, F\}$
X-ray	Binary	$\{pos, neg\}$

P	S	$P(C = T \mid P, S)$
H	T	0.05
H	F	0.02
L	T	0.03
L	F	0.001

C	$P(X = pos \mid C)$
T	0.90
F	0.20

C	$P(D = T \mid C)$
T	0.65
F	0.30

- $P(P = L) = 0.90$, $P(S = T) = 0.30$

Bayesian Artificial Intelligence

- Boolean networks a variable with n parents requires 2^{n+1} probabilities
- modeling with Bayesian networks requires the assumption of the Markov property
- there are no direct dependencies in the system being modeled which are not already explicitly shown via arcs
- in Cancer case, for example, there is no way for smoking to influence dyspnoea except by way of causing cancer (or not)
- there is no hidden "backdoor" from smoking to dyspnoea
- Bayesian networks which have the Markov property are also called Independence-maps, I-maps
- every independence suggested by the lack of an arc is real in the system

Bayesian Artificial Intelligence

- the independencies suggested by a lack of arcs are generally required to exist in the system being modeled
- it is not generally required that the arcs in a BN correspond to real dependencies in the system
- the conditional probability tables (CPTs) may be parameterized in such a way as to nullify any dependence
- for example, every fully-connected Bayesian network can represent, perhaps in a wasteful fashion, any joint probability distribution over the variables being modeled
- prefer minimal models and, in particular, minimal I-maps, which are I-maps such that the deletion of any arc violates I-mapness by implying a non-existent independence in the system

Bayesian Artificial Intelligence

- if every arc in a BN happens to correspond to a direct dependence in the system, then the BN is said to be a Dependence-map, D-map
- a BN which is both an I-map and a D-map is said to be a perfect map
- Reasoning with Bayesian networks
- how to use the Bayesian network to reason about the domain
- in particular, when we observe the value of some variable, we would like to condition upon the new information
- the process of conditioning (also called probability propagation or inference or belief updating) is performed via a "flow of information" through the network

Bayesian Artificial Intelligence

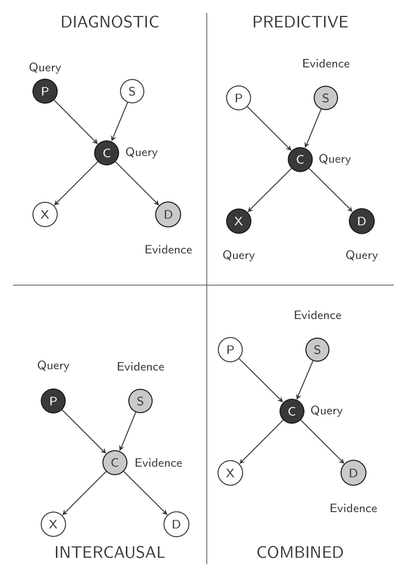
- this information flow is not limited to the directions of the arcs
- in probabilistic system, this becomes the task of computing the posterior probability distribution for a set of query nodes, given values for some evidence (or observation) nodes
- Types of reasoning
- Bayesian networks provide full representations of probability distributions over their variables
- that implies that they can be conditioned upon any subset of their variables, supporting any direction of reasoning
- for example, one can perform diagnostic reasoning, i.e., reasoning from symptoms to cause, such as

Bayesian Artificial Intelligence

- when a doctor observes Dyspnoea and then updates his belief about Cancer and whether the patient is a Smoker
- this reasoning occurs in the opposite direction to the network arcs
- one can perform predictive reasoning, reasoning from new information about causes to new beliefs about effects, following the directions of the network arcs
- for example, the patient may tell his physician that he is a smoker;
- even before any symptoms have been assessed, the physician knows this will increase the chances of the patient having cancer
- it will also change the physician's expectations that the patient will exhibit other symptoms, such as shortness of breath or having a positive X-ray result

Bayesian Artificial Intelligence

- direction of reasoning upword direction (left side)
- direction of reasoning downword direction (right side)



Bayesian Artificial Intelligence

- form of reasoning involves reasoning about the mutual causes of a common effect; this has been called intercausal reasoning
- a particular type called explaining away is of some interest
- suppose that there are exactly two possible causes of a particular effect, represented by a v-structure in the BN
- the causes Smoker and Pollution which have a common effect, Cancer
- according to the model, these two causes are independent of each other; that is,
- a patient smoking (or not) does not change the probability of the patient being subject to pollution

Bayesian Artificial Intelligence

- suppose, it is learned that Mr. Smith has cancer
- this will raise our probability for both possible causes of cancer, increasing the chances both that he is a smoker and that he has been exposed to pollution
- suppose then that it is discovered that he is a smoker
- this new information explains the observed cancer, which in turn lowers the probability that he has been exposed to high levels of pollution
- even though the two causes are initially independent, with knowledge of the effect the presence of one explanatory cause renders an alternative cause less likely

Bayesian Artificial Intelligence

- the alternative cause has been explained away
- any nodes may be query nodes and any may be evidence nodes, sometimes the reasoning does not fit neatly into one of the types described above
- the above types of reasoning can be combined in any way
- the last combination shows the simultaneous use of diagnostic and predictive reasoning
- types of evidence
- Bayesian networks can be used for calculating new beliefs when new information
- which we have been calling evidence – is available

Bayesian Artificial Intelligence

- in examples to date, we have considered evidence as a definite finding that a node X has a particular value, x which written as $X = x$
- this is sometimes referred to as specific evidence
- for example, suppose, it is discovered that the patient is a smoker, then Smoker=T, which is specific evidence
- sometimes evidence is available that is not so definite
- the evidence might be that a node Y has the value y_1 or y_2 (implying that all other values are impossible) Or
- the evidence might be that Y is not in state y_1 (but may take any of its other values); this is sometimes called a negative evidence

Bayesian Artificial Intelligence

- the new information might simply be any new probability distribution over Y
- for example, that the radiologist who has taken and analyzed the X-ray in our cancer example is uncertain
- he thinks that the X-ray looks positive, but is only 80% sure
- such information can be incorporated equivalently to Jeffrey conditionalization
- it would correspond to adopting a new posterior distribution for the node in question
- in Bayesian networks this is also known as virtual evidence
- it is handled via likelihood information, it is also known as likelihood evidence

Bayesian Artificial Intelligence

- Reasoning with numbers
- understanding Bayesian networks
- how to interpret the information encoded in a BN — the probabilistic
- semantics of Bayesian networks
- Representing the joint probability distribution
- there is a fundamental assumption that there is a useful underlying structure to the problem being modeled that can be captured with a BN, i.e.,
- that not every node is connected to every other node
- if such domain structure exists, a BN gives a more compact representation than simply describing the probability of every joint instantiation of all variables

Bayesian Artificial Intelligence

- Sparse Bayesian networks (those with relatively few arcs, which means few parents for each node) represent probability distributions in a computationally tractable way
- consider a BN containing the n nodes X_1 to X_n taken in that order
- joint distribution $P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(x_1, x_2, \ldots, x_n)$
- the chain rule

$$\begin{aligned} P(x_1, x_2, \ldots, x_n) &= P(x_1) \times P(x_2|x_1) \ldots, \times P(x_n|x_1, \ldots, x_{n-1}) \\ &= \prod_i P(x_i|x_1, \ldots, x_{i-1}) \end{aligned}$$

Bayesian Artificial Intelligence

- the structure of a BN implies that the value of a particular node is conditional only on the values of its parent nodes, this reduces to
- $P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i|Parents(X_i))$ provided $Parents(X_i) \subseteq \{x_1, \ldots, x_{i-1}\}$

$$P(X = pos \wedge D = T \wedge C = T \wedge P = low \wedge S = F)$$

$$= P(X = pos|D = T, C = T, P = low, S = F)$$

$$\times P(D = T|C = T, P = low, S = F)$$

$$\times P(C = T|P = low, S = F)P(P = low|S = F)P(S = F)$$

$$= P(X = pos|C = T)P(D = T|C = T)P(C = T|P = low, S = F)$$

$$P(P = low)P(S = F)$$

Bayesian Artificial Intelligence

- Pearl's network construction algorithm
- the condition that $Parents(X_i) \subseteq \{x_1, \ldots, x_{i-1}\}$ allows us to construct a network from a given ordering of nodes using Pearl's network construction algorithm
- the construction algorithm processes each node in order,
- adding it to the existing network and adding arcs from a minimal set of parents such that
- the parent set renders the current node conditionally independent of every other node preceding it
- using this construction algorithm, it is clear that a different node order may result in a different network structure,

Bayesian Artificial Intelligence

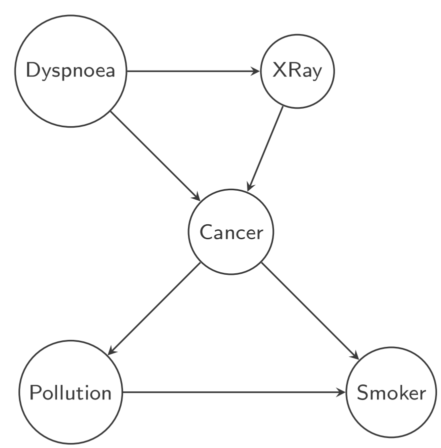
- representing the same joint probability distribution
- several different orderings will give the original network structure:
- Pollution and Smoker must be added first, but in either order, then Cancer, and then Dyspnoea and X-ray, again in either order
- on the other hand, if we add the symptoms first, we will get a markedly different network
- consider the order $< D, X, C, P, S >$. D is now the new root node
- when adding X, we must consider "Is X-ray independent of Dyspnoea?"
- since they have a common cause in Cancer, they will be dependent: learning the presence of one symptom,

Bayesian Artificial Intelligence

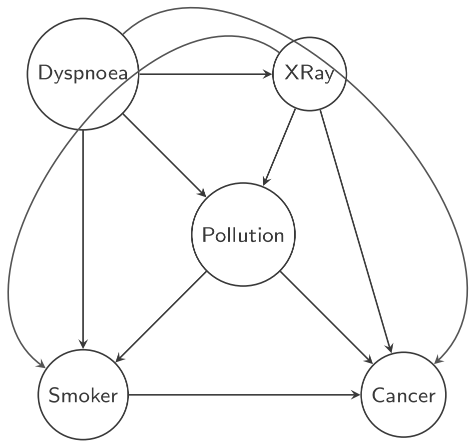
- for example, raises the probability of the other being present
- we have to add an arc from D to X, when adding Cancer,
- Cancer is directly dependent upon both Dyspnoea and X-ray, so we must add arcs from both
- for Pollution, an arc is required from C to S to carry the direct dependency
- when the final node, Smoker, is added, not only is an arc required from C to S, but another from P to S
- S and P are independent, but in the new network, without this final arc, P and S are made dependent by having a common cause, so that effect must be counter balanced by an additional arc
- the result is two additional arcs and three new probability values associated with them,

Bayesian Artificial Intelligence

- fully connected and requires as many CPT entries as a brute force specification of the full joint distribution, in such cases,
- the use of Bayesian networks offers no representational, or computational, advantage



Bayesian Artificial Intelligence



- it is desirable to build the most compact BN possible, for three reasons

Bayesian Artificial Intelligence

- the more compact the model, the more tractable it is
- it will have fewer probability values requiring specification; it will occupy less computer memory; probability updates will be more computationally efficient
- overly dense networks fail to represent independencies explicitly
- overly dense networks fail to represent the causal dependencies in the domain
- the compactness of the BN depends on getting the node ordering "right"
- the optimal order is to add the root causes first, then the variable(s) they influence directly, and continue until leaves are reached
- to understand why, we need to consider the relation between probabilistic and causal dependence

Bayesian Artificial Intelligence

- Conditional independence
- Bayesian networks which satisfy the Markov property (and so are I-maps) explicitly express conditional independencies in probability distributions
- relation between conditional independence and Bayesian network structure is important for understanding how BNs work
- consider a causal chain of three nodes, where A causes B which in turn causes C,
- $P(C|A \wedge B) = P(C|B)$
- this means that the probability of C given B, is exactly the same as the probability of C given both B and A
- knowing that A has occurred doesn't make any difference to our beliefs about C if we already know that B has occurred

Bayesian Artificial Intelligence

- the probability that someone has dyspnoea depends directly only on whether they have cancer
- if we don't know whether some woman has cancer, but we do find out she is a smoker, that would increase our belief both that she has cancer and that she suffers from shortness of breath
- if we already knew she had cancer, then her smoking wouldn't make any difference to the probability of dyspnoea
- that is, dyspnoea is conditionally independent of being a smoker given the patient has cancer
- Example statement: a new burglar alarm installed
- it reliably detects burglary, but also responds to minor earthquakes
- two neighbors, John and Mary, promise to call the police when they hear the alarm

Bayesian Artificial Intelligence

- John always calls when he hears the alarm, but sometimes confuses the alarm with the phone ringing and calls then also
- on the other hand, Mary likes loud music and sometimes doesn't hear the alarm
- given evidence about who has and hasn't called, you'd like to estimate the probability of a burglary
- BN representation of this example is shown in Figure 2.6
- all the nodes in this BN are Boolean, representing the true/false alternatives for the corresponding propositions
- this BN models the assumptions that John and Mary do not perceive a burglary directly and they do not feel minor earthquakes
- there is no explicit representation of loud music preventing Mary from hearing the alarm, nor of John's confusion of alarms and telephones;
- this information is summarized in the probabilities in the arcs from alarm to John Calls and Mary Calls

Bayesian Artificial Intelligence

- Metastatic cancer
- Example statement: Metastatic cancer is a possible cause of brain tumors and is also an explanation for increased total serum calcium
- either of these could explain a patient falling into a coma
- severe headache is also associated with brain tumors
- all the nodes are Booleans
- it is a graph, not a tree, in that there is more than one path between the two nodes M and C (via S and B)
- the basic task for any probabilistic inference system is to compute the posterior probability distribution for a set of query nodes, given values for some evidence nodes
- this task is called belief updating or probabilistic inference
- inference in Bayesian networks is very flexible, as evidence can be entered about any node while beliefs in any other nodes are updated