翻译结果

# 原文

Artificial Intelligence Week 4: Probabilities, Uncertainties, Bayesian Networks Debasis Ganguly University of Glasgow, Glasgow, UK October 14, 2024 D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 1/40  
  
Overview 1 Probabilistic Approach to address Uncertainties 2 Bayesian Networks - probability for knowledge representation and reasoning 3 Inference on Bayesian Nets D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 2/40  
  
Learning Objectives Why representation of uncertainty in agents, and why probability theory helps. Understand the basics of probability theory including sample space, events, joint distributions, conditionals. Understand, explain and be able to apply Bayes Theorem to various problems. Understand and explain how Bayesian networks can represent knowledge. Explain and apply exact inference for simple networks. Appreciate the role of sampling methods for approximate inference. D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 3/40  
  
Partial Observability and Uncertain Outcomes Have studied how rational agents should function for: Fully observable (full access to state-space at anytime, noise free) Known (environment and rules) Deterministic outcomes (no uncertainty about where you end up after taking an action) Static (e.g. the world does not change while we think about what to do next). D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 4/40  
  
Logic-based Approach to Tackle Uncertainty Example: An agent that drives taxi to the airport. ▶ : Agent decides to leave your home minutes before departure. ▶ Will ensure that you catch you flight? Issues with rule-based planning: ▶ Yes, will get there in time if no flat tire, no icy roads, no other 90 traffic, enough fuel. ▶ But are all these really Boolean (0/1) attributes? ▶ Too many rules for practical decision making! D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 5/40  
  
Rational Decision in Stochastic Environment What about ? Almost certainly gets in time but you waste lot 180 of time in the airport. Rational decision making: ▶ Relative importance of various goals - Utility Theory (Next week). ▶ Likelihood that they will be achieved - Probability Theory (Now). D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 6/40  
  
Summarizing Uncertainty Logic Rules: ▶ Enumerate all possibilities ▶ Toothache =⇒ Cavity (not always correct!) ▶ Toothache =⇒ Cavity or Gum Problem or ... ▶ Cavity =⇒ Toothache (not always correct either!) ▶ Make (contingency) plan for ALL possible eventualities (e.g. all possible sensor outcomes → grows arbitrarily large) Need more generic framework than rules connected by logic operators. Need language to describe and represent uncertainty of belief state D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 7/40  
  
Probability Describes the degree of belief in current state of the world (possibly as explained by the evidence) For example: ▶ will get me there with probability 0.85 (given all the known and 90 unknown factors in the world) ▶ If patient has toothache there is 0.8 probability that he has cavity (e.g. based on previous experience) D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 8/40  
  
Probability Introduction An experiment (or trial) is an occurrence with an uncertain outcome. ▶ E.g., we don’t know the outcome before rolling dice. ▶ For example: the outcome of dice throw is 2. Sample space: set Ω specifies all possible world states (exhaustively enumerates all possible worlds states). ▶ For dice there are 6 atomic events / sample points Ω={ , , , , , }. ω is sample point /atomic event in Ω. ▶ E.g. . D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 9/40  
  
Classical Definition of Probability Probability of an atomic event is the number of times we observe (e.g., outcome of an experiment) from the sample space out of the total number of outcomes in the sample space. ▶ P(A)=n /n Not always easy to count these (one must be careful! It’s easy to make mistakes). Look at the following example: After rolling two die, find the probability that the sum is 7. ▶ Total number of sums (n) = {2,...,12}. P(7)=1/11 (Correct?) ▶ Count the favourable pairs - {( , ),( , ),( , )}; P(7)=6/36=1/6 (Correct?). D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 10/40  
  
Axiomatic Definition of Probability P(A) > 0, where ⊂ Ω P(A) < 1 ∀A ⊂ Ω P(A) = 0 if = ∅ P(A∪B) = P(A)+P(B) if and are mutually exclusive events With the above axioms you can derive that: P(AB) = P(A)P(B) if and are independent. D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 11/40  
  
A: Event that Linda is bank teller. B: Event that Linda active in feminist movement. and are independent? Yes. P(AB)=P(A)P(B)<P(A). So, option 1 is more probable. This is an example of stereotype bias in humans. Probabilistic Thinking isn’t Natural to Humans! Linda Problem Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which one is more probable? 1 Linda is bank teller. 2 Linda is bank teller and is active in the feminist movement. D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 12/40  
  
Probabilistic Thinking isn’t Natural to Humans! Linda Problem Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which one is more probable? 1 Linda is bank teller. 2 Linda is bank teller and is active in the feminist movement. A: Event that Linda is bank teller. B: Event that Linda active in feminist movement. and are independent? Yes. P(AB)=P(A)P(B)<P(A). So, option 1 is more probable. This is an example of stereotype bias in humans. D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 12/40  
  
‘2’ can’t be more likely than ‘1’ because the probability of observing particular sequence of length 6 has to be less than that of length of 5. Work out the probabilities of the sequence yourselves as an exercise. Probabilistic Thinking isn’t Natural to Humans! Consider bin: {•, •, •, •, •, •} Consider the results of the following three “sampling with replacement” trials. Which one is more likely? 1 •, •, •, •, • 2 •, •, •, •, •, • 3 •, •, •, •, •, • D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 13/40  
  
Probabilistic Thinking isn’t Natural to Humans! Consider bin: {•, •, •, •, •, •} Consider the results of the following three “sampling with replacement” trials. Which one is more likely? 1 •, •, •, •, • 2 •, •, •, •, •, • 3 •, •, •, •, •, • ‘2’ can’t be more likely than ‘1’ because the probability of observing particular sequence of length 6 has to be less than that of length of 5. Work out the probabilities of the sequence yourselves as an exercise. D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 13/40  
  
Sample Numerical Problem box contains white balls and black balls. If balls are drawn at random without replacement, find the probability of seeing white ball by the k-th draw. = White ball is drawn by the k-th draw. {◦},{•,◦},{•,•,◦}...,{•,...,•,◦}. = {iblack balls followed by white ball is drawn} W = ∪X ∪X – These are all mutually exclusive events. 0 1 k−1 By probability axiom: P(W ) = (cid:80)k−1P(X ) i=0 P(X ) = m/(m+n); P(X ) = n/(m+n)×m/(m+n−1) and so on. 0 1 D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 14/40  
  
Conditional Probability Probability that is observed given that P(AM) is already observed: P(A|M) = . P(M) ▶ Example: P( |an even number is seen)= P( )/P(even)= 1/1 =1/3. 6 2 If ⊂ M, P(A|M) ≥ P(A) (Why?) Probability axioms hold true for any conditional M. ▶ P(A|M)>0 ▶ P(S|M)=1 (M ⊂S) ▶ P(A∪B|M)= P(AM)+P(BM) P(M) D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 15/40  
  
Numerical Example box contains 3 white balls {w ,w ,w } and 2 red balls {r ,r }. What is 1 2 3 1 2 the probability that white ball gets removed before red one? Solution w/o conditional probabilities: ▶ Space of all ordered pairs: (w ,w ), (w ,r ) and so on. 1 2 1 1 ▶ #pairs = 5×4=20 Why? ▶ Favourable pairs = 6/20 = 3/10. Solution w/ conditional probabilities (more elegant): ▶ P(W )=3/5 (event: white ball first). 1 ▶ P(R |W )=2/4 2 1 ▶ P(W )=P(R |W )×P(W )=2/4×3/5=3/10 1 2 2 1 1 D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 16/40  
  
Bayes Theorem (Overview) summarized view: P(cause|effect) ∝ P(effect|cause)P(cause) Used to estimate the probabilities from the causal direction (aka priors) to the diagnostic direction (aka posteriors). Note that the posterior is function of two different types of priors - one conditional: P(effect|cause) and the other unconditional: P(cause). The conditional needs to be look into associations of effects and causes from the past data. D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 17/40  
  
Bayes Theorem (More Formal Description) - Hypotheses/Causes (forms i partition over the set of all possibilities) - Evidence/Effect, i.e., one that is observed. Bayes Theorem - The most likely hypothesis that has led to this observation. P(A |B) = P(B|Ai)P(Ai) P(B) = P(B|Ai)P(Ai) (cid:80) P(B|Aj)P(Aj) D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 18/40  
  
A Visual Illustration Two identical looking bins: (•, •, •, •, •), and (•, •, •, •, •) You are blind-folded and asked to select ball from bin (you don’t know which bin that is). Question: You observe • ball. What is the likelihood that it came from bin B? Compute the priors: ▶ P(•|A)=3/5, P(•|A)=2/5, P(•|B)=1/5, P(•|B)=4/5 ▶ P(A) = P(B) = 1/2 (No other information is given) ▶ P(B|•) = P(•|B)P(B)/P(•) ▶ = P(•|B)P(B)/ (P(•|A)P(A) + P(•|B)P(B)) ▶ = 1/5×1/2 = 1/10 =1/4. 3/5×1/2+1/5×1/2 3/10+1/10 D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 19/40  
  
Another numerical problem The curious case of the cab - common psychological test Acabwasinvolvedinanaccident. Twocabcompanies-the•andthe•operateinthecity withg andb cabsrespectively. Awitnessidentifiedthecabinvolvedintheaccidentas• (chanceoferrorinthetestimonyduetopoorlightconditionsisα). Probabilitythatthecab involvedintheaccidentwas•? Most people just go with the witness guessing that the probability is close to 1−α. We consider the following random variables. ▶ ∈{G,B}: true color of the cab that was involved in the accident. ▶ ∈{G,B}: observed color by the witness. Need to compute P(C =G|O =B)=P(O =B|C =G)P(C = G)/(P(O =B|C =G)P(C =G)+P(O =B|C =B)P(C =B)). P(C =G|O =B)= g/(g+b)α g/(g+b)α+b/(g+b)(1−α) What happens when increases? What happens when α decreases? D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 20/40  
  
Inference by Enumeration D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 21/40  
  
Inference by Enumeration (Working example) D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 22/40  
  
Inference by Enumeration (Working example) D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 23/40  
  
Inference by Enumeration (Working example) D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 24/40  
  
Inference by Enumeration (Working example) D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 25/40  
  
Bayesian Network simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions. Syntax: ▶ set of nodes, one per random variable directed, acyclic graph (link → “directly influences”) ▶ conditional distribution for each node given its parents: P(X|Parents(X)) i ▶ In the simplest case, conditional distribution represented as conditional probability table (CPT) giving the distribution over for each possible state of the parent variables. D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 26/40  
  
Bayesian Network The topology of the network encodes conditional independence assertions. Weather is independent of the other variables. Toothache and Catch are conditionally independent given Cavity P(Toothache, Catch, Cavity, Weather)= P(Toothache|Cavity)P(Catch|Cavity) P(Cavity)P(Weather) D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 27/40  
  
Burglary Example by Judea Pearl ProblemStatement: I’matwork, neighborJohncallstosaymyalarmis ringing,butneighborMarydoesn’tcall. Sometimesthealarmissetoffbyaminor earthquake. Isthereaburglar? IdentifytheVariables: Burglar, Earthquake,Alarm,JohnCalls,MaryCalls Networktopologyreflects“causal” knowledge: ▶ Aburglarcansetthealarmoff (rareevent) ▶ Anearthquakecansetthealarmoff (rareevent) ▶ ThealarmcancauseMarytocall you(notveryreliable) ▶ ThealarmcancauseJohntocall you(fairlyreliable) D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 28/40  
  
Burglary Example (Contd.) Each node is conditionally independent of its nondescendants given its parents. is independent of b, e, and given the value of a. Full joint distribution as the product of the local conditional distributions: (cid:89) P(x ,...,x ) = P(x |Parents(X )) 1 i i=1 E.g. John calls, Mary calls, Alarm sounded, but no Burglar and no Earthquake P(j ∧m∧a∧¬b∧¬e) = P(j|a)P(m|a)P(a|¬b,¬e)P(¬b)P(¬e) = 0.9×0.7×0.001×0.999×0.998 ≈ 0.00063 (1) D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 29/40  
  
Bayesian Network Construction Construct the network such that series of locally testable assertions of conditional independence guarantees the required global semantics. Nodes: Choose an ordering of variables ,...,X 1 Any will do but more compact if the causes precedes effects. Connections: ▶ For =1...n ▶ Add to the network ▶ Select minimal set of parents from ,...,X such that 1 i−1 P(X|Parents(X))=P(X|X ,...,X ) i 1 i−1 ▶ Add link from parent(s) to i ▶ Write down the CPT such that P(X|Parents(X)) i D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 30/40  
  
Bayesian Network Construction Suppose we choose the ordering M, J, A, B, Step 1: Add MaryCalls (no parents) Step 2: Add JohnCalls ▶ Check P(J|M)=P(J)? No ▶ If Mary calls then it is likely that the alarm has gone off and John will also call. Step 3: Add Alarm ▶ P(A|J,M)=P(A|M)? No ▶ P(A|J,M)=P(A|J)? No ▶ P(A|J,M)=P(A)? No ▶ If both call, it is more likely that the alarm has gone off than if just one or neither calls, so we need both MaryCalls and JohnCalls as parents. Step 3: Add Burglary D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 31/40  
  
Bayesian Network Construction Step 4: Add Burglary (no parents) ▶ P(B|A,J,M)=P(B|A)? Yes ▶ Know the alarm state →, call from John or Mary → our phone ringing or Mary’s music, but not about burglary, → only Alarm as parent. ▶ P(B|A,J,M)=P(B)? No ▶ Alarm gives us information about whether there is burglary. Step 5: Add Earthquake ▶ P(E|B,A,J,M)=P(E|A)? No ▶ Alarm ON → more likely indicating an earthquake. ▶ If we know that there has been burglary, then that explains the alarm. ▶ This implies the probability of an earthquake would only be slightly above normal. ▶ Hence, we need both Alarm and Burglary as parents. D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 32/40  
  
Different ordering leads to different BNs Left: Causal model. Easier to explain the arrows. Middle: Diagnostic model. More dependencies introduced, e.g., the arrow between Burglary and Earthquake. Right: Bad node ordering → Yet more complex and ‘difficult to explain’ model. They all represent the same joint distribution. D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 33/40  
  
Inference on Bayesian Nets Types of inference methods Exact inference by enumeration Exact inference by variable elimination [not examinable] Approximate inference by stochastic simulation [not examinable] Approximate inference by Markov chain Monte Carlo [Types of inference methods] D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 34/40  
  
Naive Enumeration Use the basic rules of probability/Bayes and sum across relevant elements. P(B,j,m) P(B|j,m)= P(j,m) =αP(B,j,m) (2) (cid:88)(cid:88) =α P(B,e,a,j,m) a D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 35/40  
  
Naive Enumeration Rewrite full joint entries using product of CPT entries: (cid:88)(cid:88) P(B|j,m)=α P(B)P(e)P(a|B,e)P(j|a)P(m|a) a (3) (cid:88) (cid:88) =αP(B) P(e) P(a|B,e)P(j|a)P(m|a) a Exact: Yes Space: O(n) Timecomplexity: O(2n)foraBoolean network Ingeneral: polynomialtimeongeneral trees(NP-hardongeneralgraphs) Issue: Inefficient,sincerepeated computatione.g.,computes P(j|a)P(m|a)foreachvalueofa. D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 36/40  
  
Sampling-based Sample from P(Cloudy)=<0.5,0.5>, value is true. Sample from P(Sprinkler|Cloudy =true)=< 0.1,0.5>, value is false. Sample from P(Rain|Cloudy = true)=<0.8,0.2>, value is true. Sample from P(WetGrass|Sprinkler = false,Rain=true)=<0.9,0.1>, value is true. Sampled events [true,false,true,true]. D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 37/40  
  
Sampling-based Probability that the procedure generates particular event (x ,...,x )= PS 1 (cid:81)n P(x|parents(X))=P(x ,...,x ) i=1 i 1 In general, let (x ,...,x ) be the number PS 1 of samples generated for event ,...,x 1 lim P′(x ...x )= lim (x ,...,x )/N 1 PS 1 N→∞ N→∞ =S (x ,...,x ) (4) PS 1 =P(x ...x ) 1 i.e., P′(x ,...,x )≈P(x ...,x ). 1 1 Issue: Need huge number of samples. D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 38/40  
  
Approximate Sampling for Faster Inference Naive Sampling takes massive number of sampling steps. Approximate Inference via Gibbs Sampling or MCMC can reduce computation time. Basic Idea of Gibbs Sampling: ▶ Generate next state by sampling one variable given its Markov blanket. ▶ Sample each variable in turn, keeping evidence fixed. ▶ Estimate P(Rain|Sprinkler =true,WetGrass =true) For our example: ▶ Sample Cloudy or Rain given its Markov blanket (parents + children + children’s parents), repeat. ▶ Count number of times Rain is true and false in the samples. ▶ E.g., visit 100 states: 31 have Rain=true, 69 have Rain=false ▶ P(Rain|Sprinkler =true, WetGrass =true)=Normalize(<31,69> )=<0.31,0.69> D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 39/40  
  
Summary Probabilistic reasoning: advantages over logical reasoning when there is not enough information to be sure actions will work. Belief networks / Bayesian Networks ▶ Data structures for representing dependence among variables ▶ joint probability distribution ▶ cause-effect relationships ▶ Inference: computing the p.d.f. of subset of variables, given set of evidence variables. Next Week: ▶ Study Utility Theory ▶ Combine utility with probabilistic reasoning for decision making under uncertainty. D.Ganguly ArtificialIntelligenceWeek4:Probabilities,UncertainOtiecst,obBearye1s4i,an20N2e4tworks 40/40

# 翻译

人工智能 第四周：概率、不确定性、贝叶斯网络  
Debasis Ganguly 格拉斯哥大学，英国格拉斯哥 2024年10月14日 D. Ganguly 人工智能第四周：概率、不确定性、贝叶斯网络 1/40  
  
概览  
1. 概率方法处理不确定性  
2. 贝叶斯网络——用于知识表示和推理的概率  
3. 贝叶斯网络上的推理 D. Ganguly 人工智能第四周：概率、不确定性、贝叶斯网络 2/40  
  
学习目标  
- 为什么在智能体中表示不确定性，以及概率论如何帮助解决这一问题。  
- 理解概率论的基础知识，包括样本空间、事件、联合分布、条件概率。  
- 理解、解释并能够将贝叶斯定理应用于各种问题。  
- 理解并解释贝叶斯网络如何表示知识。  
- 解释并应用精确推理于简单网络。  
- 认识到采样方法在近似推理中的作用。  
  
Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 3/40  
  
部分可观测性和不确定的结果 已经研究了理性代理在以下情况下的工作方式：完全可观测（随时可以无噪声地访问状态空间）已知（环境和规则）确定性结果（采取行动后对最终位置没有不确定性）静态（例如，在我们考虑下一步做什么时，世界不会发生变化）。D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 4/40  
  
基于逻辑的方法处理不确定性 示例：一个驾驶出租车前往机场的代理。 ▶ 代理决定在出发前几分钟离开你的家。 ▶ 将确保你赶上航班。 基于规则规划的问题： ▶ 是的，如果轮胎没有漏气、道路不结冰、没有其他交通堵塞、有足够的燃料，那么确实能够准时到达。 ▶ 但是这些真的都是布尔（0/1）属性吗？ ▶ 实际决策中需要太多的规则。 D  
  
Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 5/40  
  
随机环境中的理性决策 如果这样的话。几乎可以肯定能按时到达，但你会在机场浪费很多时间。理性决策： ▶ 各种目标的相对重要性 - 效用理论（下周）。 ▶ 实现这些目标的可能性 - 概率论（现在）。 D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 6/40  
  
总结不确定性 逻辑规则： ▶ 列出所有可能性 ▶ 牙疼 =⇒ 蛀牙（并不总是正确的。） ▶ 牙疼 =⇒ 蛀牙或牙龈问题或... ▶ 蛀牙 =⇒ 牙疼（这也不总是正确的。） ▶ 为所有可能的情况制定（应急）计划（例如，所有可能的传感器结果 → 可以任意增长） 需要比通过逻辑运算符连接的规则更通用的框架。需要一种语言来描述和表示信念状态的不确定性 D  
  
Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 7/40  
  
概率 描述了对当前世界状态（可能由证据解释）的置信程度。例如： ▶ 我将以0.85的概率到达那里（基于世界上所有已知和未知因素）。 ▶ 如果患者有牙痛，那么他有蛀牙的概率为0.8（例如，基于以往的经验）。 D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 8/40  
  
概率 介绍 一个实验（或试验）是一个具有不确定结果的发生事件。 ▶ 例如，我们在掷骰子之前不知道其结果。 ▶ 例如：掷骰子的结果是2。 样本空间：集合Ω指定了所有可能的世界状态（详尽地列举了所有可能的世界状态）。 ▶ 对于骰子来说，有6个原子事件/样本点 Ω={1, 2, 3, 4, 5, 6}。ω 是 Ω 中的样本点/原子事件。 ▶ 例如 D  
  
Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 9/40  
  
概率的经典定义 原子事件的概率是从样本空间中观察到（例如，实验结果）的次数除以样本空间中的总结果数。▶ P(A)=n /n 并不总是容易计算这些（必须小心。很容易出错）。看下面的例子：掷两个骰子后，找到和为7的概率。▶ 总和的数量（n）= {2, ..., 12}。P(7)=1/11（正确。） ▶ 计算有利的组合 - {( , ), ( , ), ( , )}; P(7)=6/36=1/6（正确。）。D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 10/40  
  
概率的公理化定义 P(A) > 0，其中 ⊂ Ω P(A) ≤ 1 ∀A ⊂ Ω 如果 = ∅ 则 P(A) = 0 如果 和 是互斥事件，则 P(A∪B) = P(A)+P(B) 根据上述公理可以推导出：如果 和 是独立事件，则 P(AB) = P(A)P(B) D  
  
Ganguly 人工智能第四周：概率、不确定性、贝叶斯网络 11/40  
  
A：琳达是银行出纳员的事件。B：琳达积极参与女权运动的事件。A和B是独立事件。是的。P(AB)=P(A)P(B)<P(A)。因此，选项1更有可能。这是人类刻板印象偏差的一个例子。概率思维对人类来说并不自然。琳达问题 琳达31岁，单身，直言不讳，非常聪明。她主修哲学。作为学生，她非常关注歧视和社会正义问题，并且还参加了反核示威活动。哪一个更有可能？1. 琳达是银行出纳员。2. 琳达是银行出纳员并且积极参与女权运动。D. Ganguly 人工智能第四周：概率、不确定性、贝叶斯网络 12/40  
  
概率思维对人类来说并不自然。琳达问题 琳达31岁，单身，直言不讳，非常聪明。她主修哲学。  
  
作为一名学生，她非常关注歧视和社会正义问题，并且还参加了反核示威。哪一个更有可能：1. 琳达是银行出纳员。2. 琳达是银行出纳员并且积极参与女权运动。A: 事件琳达是银行出纳员。B: 事件琳达积极参与女权运动。这两个事件是独立的。是的。P(AB)=P(A)P(B)<P(A)。因此，选项1更有可能。这是人类刻板印象偏见的一个例子。D. Ganguly 人工智能第四周：概率、不确定性、贝叶斯网络 12/40  
  
‘2’不可能比‘1’更可能，因为观察到长度为6的特定序列的概率必须小于长度为5的序列的概率。作为练习，请自行计算这些序列的概率。概率思维对人类来说并不自然。考虑一个箱子：{•, •, •, •, •, •} 考虑以下三个“有放回抽样”试验的结果。哪一个更有可能  
  
1 •, •, •, •, • 2 •, •, •, •, •, • 3 •, •, •, •, •, • D. Ganguly 人工智能第四周：概率、不确定性、贝叶斯网络 13/40  
  
概率思维对人类来说并不自然。考虑一个箱子：{•, •, •, •, •, •} 考虑以下三个“有放回抽样”试验的结果。哪一个更有可能发生。  
1 •, •, •, •, •  
2 •, •, •, •, •, •  
3 •, •, •, •, •, •  
‘2’不可能比‘1’更有可能，因为观察到长度为6的特定序列的概率必须小于长度为5的序列的概率。作为练习，请自行计算这些序列的概率。  
D. Ganguly 人工智能第四周：概率、不确定性、贝叶斯网络 13/40  
  
示例数值问题  
一个箱子里装有白色球和黑色球。如果从中随机无放回地抽取球，请找出在第k次抽取时看到白球的概率。  
= 在第k次抽取时抽到白球。  
{◦},{•,◦},{•,•,◦},...,{•,...,•,◦}  
  
= {先抽取i个黑球后抽取一个白球} W = ∪X ∪X – 这些都是互斥事件。0 1 k−1 根据概率公理：P(W) = (cid:80)k−1P(X ) i=0 P(X ) = m/(m+n); P(X ) = n/(m+n)×m/(m+n−1)，以此类推。0 1 D. Ganguly 人工智能第4周：概率、不确定性、贝叶斯网络 14/40  
  
条件概率 已知观察到M的情况下，观察到A的概率：P(A|M) = P(AM)/P(M) ▶ 例子：P( | 观察到偶数)= P( )/P(偶数)= 1/6 / 1/2 =1/3。如果 ⊂ M, P(A|M) ≥ P(A)（为什么？）对于任何条件下的M，概率公理都成立。 ▶ P(A|M)>0 ▶ P(S|M)=1 （M ⊂S） ▶ P(A∪B|M)= P(AM)+P(BM)/P(M) D. Ganguly 人工智能第4周：概率、不确定性、贝叶斯网络 15/40  
  
数值示例 一个盒子中装有3个白球{w , w , w }和2个红球{r , r }。求在红球之前取出白球的概率。1 2 3 1 2 不使用条件概率的解法： ▶ 所有序对的空间：(w , w ), (w , r ) 等等1 2 1 2  
  
1 2 1 1 ▶ #对数 = 5×4=20 为什么。▶ 有利的对数 = 6/20 = 3/10。使用条件概率的解法（更优雅）：▶ P(W )=3/5 （事件：第一次抽到白球）。1 ▶ P(R |W )=2/4 2 1 ▶ P(W )=P(R |W )×P(W )=2/4×3/5=3/10 1 2 2 1 1 D. Ganguly 人工智能第4周：概率、不确定性、贝叶斯推理与网络 16/40  
  
贝叶斯定理（概述）简化视图：P(原因|结果) ∝ P(结果|原因)P(原因) 用于从因果方向（也称为先验）估计概率至诊断方向（也称为后验）。请注意，后验是两种不同类型先验的函数——一种是有条件的：P(结果|原因)，另一种是无条件的：P(原因)。有条件的那种需要查看过去数据中结果与原因之间的关联。D  
  
Ganguly 人工智能 第四周：概率、不确定性、贝叶斯和网络 17/40  
  
贝叶斯定理（更正式的描述）- 假设/原因（在所有可能性集合上形成一个划分）- 证据/结果，即被观察到的事物。贝叶斯定理 - 导致这一观察结果最可能的假设。  
\[P(A|B) = \frac{P(B|A\_i)P(A\_i)}{P(B)} = \frac{P(B|A\_i)P(A\_i)}{\sum P(B|A\_j)P(A\_j)}\]  
D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯和网络 18/40  
  
一个视觉说明  
两个外观相同的箱子：(•, •, •, •, •)，以及 (•, •, •, •, •)。你被蒙住眼睛并要求从其中一个箱子里选取一个球（你不知道是哪个箱子）。问题：你观察到了一个•球。它来自箱子B的可能性有多大？  
计算先验概率：  
- \(P(•|A)=3/5\)，\(P(•|A)=2/5\)，\(P(•|B)=1/5\)，\(P(•|B)=4/5\)  
- \(P(A) = P(B) = 1/2\)（没有给出其他信息）  
- \(P(B|•) = \frac{P(•|B)P(B)}{P(•)}\)  
- \(= \frac{P(•|B)P(B)}{(P(•|A)P(A) + P(•|B)P(B))}\)  
- \(= \frac{1/5×1/2}{(3/5×1/2 + 1/5×1/2)}\)  
- \(= \frac{1/10}{1/4} = 1/4\)  
  
3/5×1/2+1/5×1/2 3/10+1/10 D. Ganguly 人工智能第四周：概率、不确定性、贝叶斯网络 19/40  
  
另一个数值问题 出租车的奇怪案例 - 常见的心理测试 一辆出租车卷入了一起事故。城市中有两家出租车公司——绿色和蓝色，分别运营着g辆和b辆出租车。一位目击者在光线条件不佳的情况下（错误识别的概率为α）将事故中的出租车辨认为蓝色。求事故中涉及的出租车是蓝色的概率。大多数人只是根据目击者的说法猜测这个概率接近于1−α。我们考虑以下随机变量。  
- C ∈ {G, B}：实际卷入事故的出租车的真实颜色。  
- O ∈ {G, B}：目击者观察到的颜色。  
  
需要计算P(C = G | O = B) = P(O = B | C = G)P(C = G) / (P(O = B | C = G)P(C = G) + P(O = B | C = B)P(C = B))。  
\[P(C = G | O = B) = \frac{\frac{g}{g+b}\alpha}{\frac{g}{g+b}\alpha + \frac{b}{g+b}(1-\alpha)}\]  
  
当g增加时会发生什么？当α减少时又会发生什么？  
  
Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 20/40  
  
枚举推理 D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 21/40  
  
枚举推理（工作示例） D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 22/40  
  
枚举推理（工作示例） D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 23/40  
  
枚举推理（工作示例） D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 24/40  
  
枚举推理（工作示例） D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 25/40  
  
贝叶斯网络是一种简单、图形化的表示方法，用于条件独立性断言，因此可以紧凑地指定完整的联合分布。  
  
语法：▶ 一组节点，每个随机变量一个节点，有向无环图（链接 → “直接影响”） ▶ 给定其父节点的每个节点的条件分布：P(X|Parents(X)) i ▶ 在最简单的情况下，条件分布表示为条件概率表 (CPT)，给出对于每个可能的父变量状态下的分布。D. Ganguly 人工智能第四周：概率、不确定性、贝叶斯网络 26/40  
  
贝叶斯网络 网络的拓扑结构编码了条件独立性断言。天气与其他变量独立。牙痛和捕获在给定蛀牙的情况下是条件独立的 P(牙痛, 捕获, 蛀牙, 天气) = P(牙痛|蛀牙)P(捕获|蛀牙)P(蛀牙)P(天气) D. Ganguly 人工智能第四周：概率、不确定性、贝叶斯网络 27/40  
  
入室盗窃示例 由Judea Pearl提出 问题陈述：我在工作，邻居John打电话来说我的警报响了，但邻居Mary没有打电话。  
  
有时候警报是由轻微的地震触发的。是小偷吗？识别变量：小偷、地震、警报、约翰打电话、玛丽打电话。网络拓扑反映了“因果”知识：  
- 小偷可以触发警报（罕见事件）  
- 地震可以触发警报（罕见事件）  
- 警报可能导致玛丽给你打电话（不太可靠）  
- 警报可能导致约翰给你打电话（相当可靠）  
D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 28/40  
  
入室盗窃示例（续）每个节点在其父节点给定的情况下，与其非后代节点条件独立。在给定a值的情况下，与b、e和独立。全联合分布作为局部条件分布的乘积：  
\[ \prod P(x\_1, ..., x\_n) = P(x\_i | Parents(X\_i)) \]  
例如，约翰打电话，玛丽打电话，警报响起，但没有小偷也没有地震  
\[ P(j \land m \land a \land \neg b \land \neg e) = P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e) \]  
\[ = 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 \approx 0.00063 \] (1)  
D  
  
Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 29/40  
  
贝叶斯网络构建  
构建网络，使得一系列局部可测试的条件独立断言能够保证所需的全局语义。  
节点：  
选择变量的顺序 ,. ,X1。任何顺序都可以，但如果原因在结果之前则更加紧凑。  
连接：  
- 对于 i=1...n  
- 将 添加到网络中  
- 从 ,. ,X1 中选择最小的父节点集合，使得 P(Xi|Parents(Xi))=P(Xi|X1,...,Xi-1)  
- 从父节点到 添加链接  
- 记录下条件概率表（CPT），使得 P(Xi|Parents(Xi))  
D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 30/40  
  
贝叶斯网络构建  
假设我们选择了 M, J, A, B 的顺序，  
步骤1：添加 MaryCalls（无父节点）  
步骤2：添加 JohnCalls  
- 检查 P(J|M)=P(J)。不成立  
- 如果玛丽打电话了，那么很可能警报已经响起，并且约翰也会打电话。  
步骤3：添加 Alarm  
- P(A|J,M)=P(A|M)  
  
No ▶ P(A|J,M)=P(A|J)。不成立 ▶ P(A|J,M)=P(A)。不成立 ▶ 如果两者都打电话，比起只有一个人或没有人打电话，警报响起的可能性更大，因此我们需要将MaryCalls和JohnCalls作为父节点。步骤3：添加Burglary D. Ganguly 人工智能第4周：概率、不确定性、贝叶斯网络 31/40  
  
贝叶斯网络构建 步骤4：添加Burglary（无父节点） ▶ P(B|A,J,M)=P(B|A)。成立 ▶ 知道警报状态 → John或Mary的电话 → 我们的电话铃声或Mary的音乐，但不知道是否发生了盗窃，→ 只有Alarm作为父节点。 ▶ P(B|A,J,M)=P(B)。不成立 ▶ 警报为我们提供了是否有盗窃的信息。步骤5：添加地震 ▶ P(E|B,A,J,M)=P(E|A)。不成立 ▶ 警报开启 → 更有可能表明发生了地震。 ▶ 如果我们知道发生了盗窃，那么这就解释了警报的原因。 ▶ 这意味着发生地震的概率只会略高于正常水平。 ▶ 因此，我们需要同时将Alarm和Burglary作为父节点。 D  
  
Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 32/40  
  
不同的排序会导致不同的贝叶斯网络。左图：因果模型。更容易解释箭头的方向。中图：诊断模型。引入了更多的依赖关系，例如，入室盗窃和地震之间的箭头。右图：不良的节点排序 → 更加复杂且“难以解释”的模型。它们都表示相同的联合分布。D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 33/40  
  
贝叶斯网络上的推理 推理方法类型  
- 通过枚举进行精确推理  
- 通过变量消元进行精确推理 [不考]  
- 通过随机模拟进行近似推理 [不考]  
- 通过马尔可夫链蒙特卡洛方法进行近似推理 [推理方法类型] D  
  
Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 34/40  
  
朴素枚举 使用基本的概率/贝叶斯规则，并对相关元素求和。 P(B,j,m) P(B|j,m)= P(j,m) =αP(B,j,m) (2) (cid:88)(cid:88) =α P(B,e,a,j,m) a D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 35/40  
  
朴素枚举 通过条件概率表（CPT）条目的乘积重写完整的联合条目： (cid:88)(cid:88) P(B|j,m)=α P(B)P(e)P(a|B,e)P(j|a)P(m|a) a (3) (cid:88) (cid:88) =αP(B) P(e) P(a|B,e)P(j|a)P(m|a) a 精确性：是 空间复杂度：O(n) 时间复杂度：对于布尔网络为O(2^n) 一般情况下：在一般树上为多项式时间（在一般图上为NP难问题） 问题：效率低下，因为重复计算，例如，对于a的每个值都计算P(j|a)P(m|a)。 D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 36/40  
  
基于采样的方法 从P(多云)=<0.5,0.5>中采样，得到值为真。 从P(洒水器|多云=真)=<0.1,0  
  
5>, 值为假。从P(Rain|Cloudy = true)=<0.8, 0.2>中采样，值为真。从P(WetGrass|Sprinkler = false, Rain=true)=<0.9, 0.1>中采样，值为真。采样事件 [true, false, true, true]。D. Ganguly 人工智能第四周：概率、不确定性、贝叶斯网络 37/40  
  
基于采样的生成特定事件 (x ,..., x ) 的概率 PS 1 (cid:81)n P(x|parents(X))=P(x ,..., x ) i=1 i 1 通常情况下，设 (x ,..., x ) 是为事件 ,..., x 1 生成的样本数量 PS 1 lim P′(x,..., x )= lim (x ,..., x )/N 1 PS 1 N→∞ N→∞ =S (x ,..., x ) (4) PS 1 =P(x,..., x ) 1 即，P′(x ,..., x )≈P(x,..., x )。 1 1 问题：需要大量的样本。D. Ganguly 人工智能第四周：概率、不确定性、贝叶斯网络 38/40  
  
近似采样以加速推理 简单采样需要大量的采样步骤。通过吉布斯采样或MCMC进行近似推理可以减少计算时间。  
  
Gibbs采样的基本思想：▶ 通过给定其马尔可夫毯来抽样一个变量以生成下一个状态。▶ 依次对每个变量进行抽样，保持证据不变。▶ 估计P(下雨|洒水器=true,草地湿润=true) 对于我们的例子：▶ 给定其马尔可夫毯（父节点+子节点+子节点的父节点）抽样多云或下雨，重复此过程。▶ 计算样本中下雨为真和假的次数。▶ 例如，访问100个状态：31次下雨=true，69次下雨=false ▶ P(下雨|洒水器=true,草地湿润=true)=归一化(<31,69>)=<0.31,0.69> D. Ganguly 人工智能第四周：概率、不确定性推理、贝叶斯网络 39/40  
  
总结 概率推理：当信息不足以确定行动是否有效时，相比逻辑推理具有优势。信念网络/贝叶斯网络 ▶ 表示变量间依赖关系的数据结构 ▶ 联合概率分布 ▶ 因果关系 ▶ 推理：计算概率密度函数  
  
在给定证据变量集的情况下，计算变量子集的概率。下周内容：▶ 学习效用理论 ▶ 结合效用来进行不确定性条件下的概率推理决策。D. Ganguly 人工智能 第四周：概率、不确定性、贝叶斯网络 2024年1月4日 40/40