Байесовы сети

Вероятностное моделирование в Байесовых сетях

Вероятностная трактовка данных

- Процесс измерения сопряжен с экспериментальными погрешностями;
- Изучаемая система является сложной, т.е. несводимой к сумме свойств отдельных компонент, и наблюдаемое многообразие данных может быть равновероятно объяснено великим множеством структурных описаний, при этом нельзя достоверно предпочесть ни одно из них;
- Объем измерений конечен и не может считаться исчерпывающим описанием системы.

Как согласовать субъективные неопределенности в ожиданиях исследователей (beliefs) и объективные факты о статистике экспериментальных наблюдений (probabilities)?

Частотные (экспериментальные) вероятности

- Предельный результат «бесконечного» числа испытаний в повторяющихся условиях;
- *Пример*:
  «Вероятность выпадения орла при бросании *этой* монеты равна 0.52»;
- Вычисляется как доля «орлов» в пределе неограниченного числа попыток.

Повторяющиеся условия? Неограниченное число попыток? Как экспериментально установить вероятность 10^{-6} ?

Байесовы вероятности (ожидания)

- Количественное выражение степени ожиданий;
- *Пример*:
  «Вероятность того, что завтра будет дождь есть 0.3»;
- «Завтра» нельзя повторить;
- Ожидания *субъективны* и зависят от (индивидуальной) априорной информации;
- Частотные распределения наблюдений являются частным случаем.

Все вероятности являются условными (обусловленными имеющейся информацией и опытом эксперта). Для получения «объективной» картины предполагается усреднение.

Субъективные ожидания

- Требуется оценить вероятность положительного исхода в каждой из трех ситуаций:

1. *Знатная леди* утверждает, что она может отличить на вкус, был ли чай налит в сливки или наоборот – сливки в чай. Ей удалось это проделать 10 раз в течение бала;
2. *Азартный игрок* утверждает, что он может предсказать, орлом или решкой выпадет монета (которую вы ему дадите). Он смог выиграть такое пари уже 10 раз за этот вечер, ни разу не проиграв;
3. *Эксперт в классической музыке* декларирует, что он в состоянии различить творения Гайдна и Моцарта лишь по одной странице партитуры. Он уверенно проделал это 10 раз в музыкальной библиотеке.

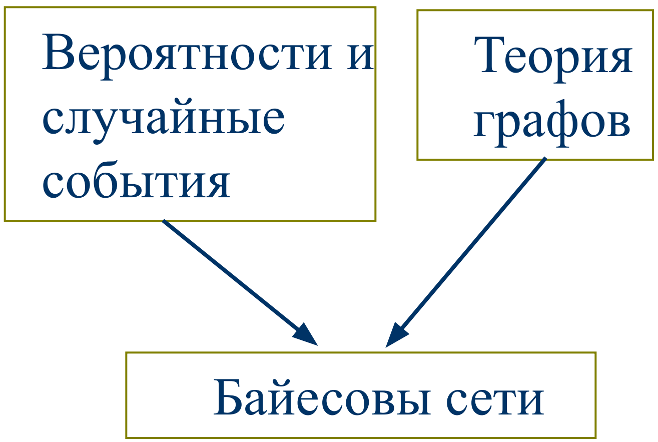
Наши субъективные оценки вероятности этих трех ситуаций весьма отличаются, хотя мы имеем дело с повторяющимися событиями!

О повторяемости условий случайных событий

- Серия экспериментов с бросанием кубика, сделанного из сахара, на влажную поверхность стола.
- Вероятности исходов последующих испытаний зависят от относительной частоты исходов предыдущих испытаний, при этом исследуемая система каждый раз необратимо изменяется в результате каждого эксперимента.

Такими свойствами обладают многие биологические, экономические и социальные системы.

Графы и вероятности: Байесовы сети



- Случайные переменные – узлы *ориентированного* графа
- Отношения прямой зависимости – ребра графа
- Каждая переменная может зависеть только от некоторого множества своих соседей
- Плотность совместной вероятности значений всех переменных редуцируется до произведения локальных условных плотностей

Применения Байесовых сетей: медицина, социология, стратегическое планирование, риски, управление, финансы и экономика

Преподобный Томас Байес

- Преп. Томас Байес, математик, впервые использовавший вероятность в индуктивном смысле, и установивший основы вероятностного рассуждения и вывода.
- Какова вероятность события в будущих испытаниях, основываясь на числе предыдущих испытаний, в которых это событие *не произошло*? Теорема Байеса связывает априорные и апостериорные вероятности причин после наблюдения следствий.
- Труд «Essay Towards Solving a Problem in the Doctrine of Chances», Philosophical Transactions of the Royal Society of London (1763).
- Томас Байес похоронен на Bunhill Fields в центре Лондона (сейчас это парк).

Bayes, Thomas (b. 1702, London – d. 1761, Tunbridge Wells, Kent),

Эксперты и экспертные системы

«Логика» действий эксперта:

- Получение информации о состоянии окружающего мира
- Принятие решения, выбор действий, по поводу которых имеются определенные ожидания последствий
- Приобретение опыта путем сопоставления результатов действий и ожиданий и возврат к первому этапу.

Развитие экспертных систем:

- От моделирования эксперта к моделированию предметной области
- От попыток учета неопределенности в правилах к использованию классической теории вероятностей и теории принятия решений
- От попыток замены эксперта – к оказанию ему помощи.

Новый опыт и информация о мире позволяют эксперту сообразно действовать в будущем

Особенности вывода суждений в условиях неопределенности

- Суть приобретаемого знания в условиях неопределенности состоит в понимании, влияет ли полученная информация на наши ожидания относительно других событий. Основная причина трудностей при использовании систем, основанных на правилах, состоит в учете "сторонних", "косвенных" последствий наблюдаемых событий;
- Сложности с интуитивным пониманием независимости: условная независимость, индуцированная зависимость, «попутное» объяснение.

Байесова логика: суждение базируется на предыдущем опыте, уточняемом *имеющимся* объемом оперативной информации.

Задача о траве Шерлока Холмса

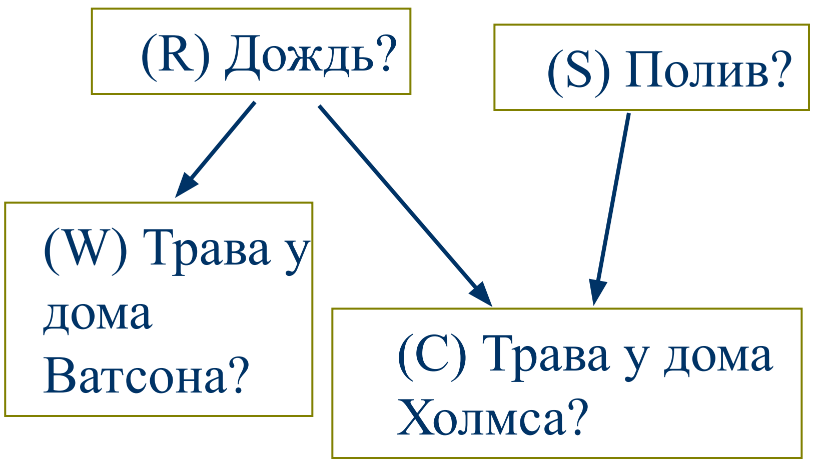
- Шерлок Холмс вышел из дома утром и заметил, что трава вокруг влажная. Он рассудил: «Я думаю, что ночью был дождь. Следовательно, трава возле дома моего соседа, доктора Ватсона, вероятно, также влажная». Таким образом, информация о состоянии травы у дома Холмса повлияла на его ожидания относительно влажности травы у дома Ватсона.
- Холмс проверил состояние сборника дождевой воды и обнаружил, что тот - сухой. В результате Холмс вынужден изменить ход своих рассуждений, и состояние травы возле его дома перестает влиять на ожидания по поводу травы у соседа.
- Помимо дождя, Холмс мог просто забыть выключить поливальную установку накануне. Допустим, на следующее утро Холмс снова обнаруживает, что трава влажная. Это повышает его субъективные вероятности и для прошедшего дождя, и по поводу забытой дождевальной установки. Затем Холмс замечает, что трава у дома Ватсона также влажная, и заключает, что ночью был дождь.

Попутное объяснение (редукция причины, explaining away)

- Влажность травы у дома Холмса объясняется дождем, и следовательно нет оснований продолжать ожидать, что была забыта включенной поливальная машина. Следовательно, возросшая, было, субъективная вероятность относительно забытой поливальной машины уменьшается до (практически) исходного значения, имевшего место до выхода Холмса из дома.

Этот шаг рассуждений практически невозможно воспроизвести в машинных системах, основанных на правилах, однако он абсолютно естественен для человека.

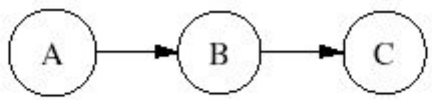
Задача о траве Шерлока Холмса



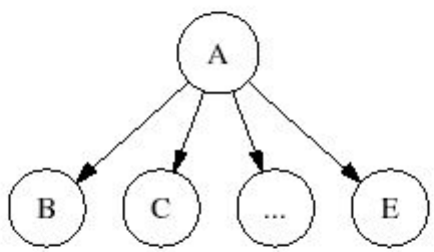
- Задача описывается Байесовой сетью с 4 переменными;
- Причины влажности травы у дома Холмса – дождь и/или поливальная установка;
- Как меняются вероятности этих причин после поступления информации о сборнике дождевой воды и о влажности травы возле дома Ватсона?

Explaining Away: Как возникает попутное объяснение?

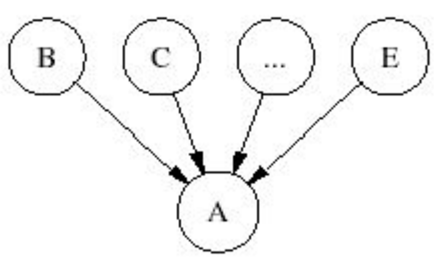
Отношения между переменными в Байесовых сетях



(a)



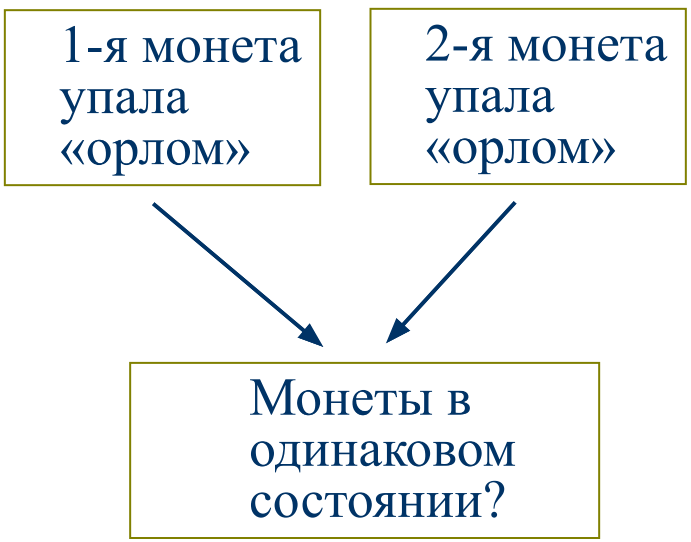
(b)



(c)

- *Последовательное соединение* (a)
  Влияние может распространяться от A к C и обратно, пока B не конкретизировано.
- *Дивергентное соединение* (b)
  Влияние может распространяться между потомками узла A, пока его значение не конкретизировано.
- *Конвергентное соединение* (c)
  Если об A ничего не известно, кроме того, что может быть выведено из информации о его предках B,C,…,E, то эти переменные предки являются разделенными. При уточнении A открывается канал взаимного влияния между его предками.

Индуцированная зависимость в конвергентных соединениях



- Две переменные соответствуют независимым подбрасываниям двух разных монет;
- Эти две переменные являются *причинами* для 3-й переменной-индикатора;
- Если известно значение 3-й переменной, то первые две переменные становятся зависимыми: знание одной полностью определяет другую.

Сложности с интуитивным пониманием независимости

- Индуцированная зависимость в конвергентных соединениях.
- Explaining away – “попутное объяснение”.
- Условная зависимость – мысленное выделение группы фактов, для которых значение обуславливающей переменной определено.
- «Парадоксы»:
  - Доход, пол и возраст в обществе
  - Умственные и физические способности студентов колледжа

Формализация понятия независимости: d-разделимость

- Определение (d-разделимость). Две переменные A и B в Байесовой сети являются d-разделенными, если на каждом пути, соединяющем эти две вершины на графе, найдется промежуточная переменная V, такая что:
  - Соединение с V последовательное или дивергентное, и значение V известно, либо
  - Соединение конвергентное, и нет свидетельств ни о значении V, ни о каждом из ее потомков.

В задаче Холмса переменные «Полив» и «Влажная трава у дома Ватсона» d-разделены, если известно, что был дождь (дивергентное соединение на графе).

Вероятности отдельных переменных

$$P(A|B) = x$$

$$P(A,B) = P(A|B) \cdot P(B)$$

$$P(A_j|B) = \frac{P(A_j) \cdot P(B|A_j)}{P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)}$$

- Условная вероятность – вероятность появления А при условии В («после» В)
- Формула полной вероятности
- Теорема Байеса

Теорема Байеса дает решение обратной задачи - какова вероятность наступления более раннего события В, если известно что более позднее событие А наступило?

Пример рассуждений на основе теоремы Байеса

$$P(H_1) = 0.9 \quad P(A|H_1) = 0.15$$

$$P(H_2) = 0.1 \quad P(A|H_2) = 0.92$$

$$P(H_1|A) = \frac{0.15 \cdot 0.9}{0.15 \cdot 0.9 + 0.92 \cdot 0.1} = 0.595$$

$$P(H_2|A) = \frac{0.92 \cdot 0.1}{0.15 \cdot 0.9 + 0.92 \cdot 0.1} = 0.405$$

- Факт: в водоеме обнаружено загрязнение с превышением ПДК.
- Потенциальные источники - два предприятия, причем выбросы на первом происходят в 9 раз чаще, чем на втором.
- Только 15% сбросов первого предприятия превышают ПДК. Для второго предприятия эта вероятность равна 92%
- Кто виноват?!

Байесова сеть

- Байесова сеть состоит из следующих понятий и компонент:
  - Множество случайных переменных и направленных связей между переменными;
  - Каждая переменная может принимать одно из конечного множества взаимоисключающих значений;
  - Переменные вместе со связями образуют ориентированный граф без циклов;
  - К каждой переменной-потомку A с переменными-предками $B_1,\ldots,B_n$ приписывается таблица условных вероятностей $P(A|\,B_1,\ldots,B_n)$

Редукция полной вероятности в Байесовой сети

- Редукция совместной вероятности распределения нескольких случайных переменных в Байесовой сети:

$$P(A_1,\ldots,A_n)=\prod_j P\left(A_j \middle| pa\left(A_j\right)\right)$$

$pa\left(A_j\right)$ - состояния всех переменных – предков для переменной A_j. Это выражение носит название цепного правила для полной вероятности.

- Пример редукции вероятности в задаче Холмса:

4D

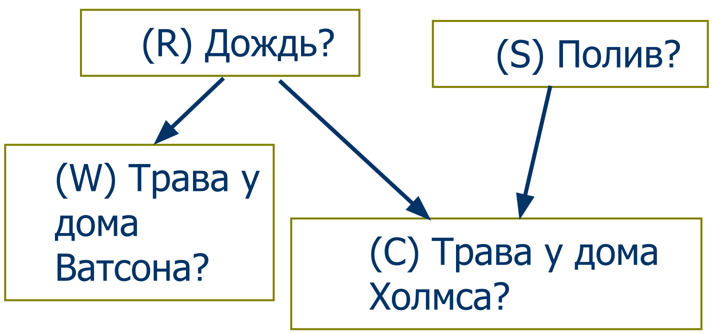
$$P(R,S,C,W)=P(R)\cdot P(S|R)\cdot P(C|R,S)\cdot P(W|R,S,C)$$

$$P(R,S,C,W)=P(R)\cdot P(S)\cdot P(C|R,S)\cdot P(W|R)$$

2D

Вычисления в задаче Холмса

Граф Байесовой сети



Таблицы вероятностей

R	P(R)
t	0.3
f	0.7

S	P(S)
t	0.2
f	0.8

R	P(W=t|R)	P(W=f|R)
t	0.8	0.2
f	0.1	0.9

R	S	P(C=t|R,S)	P(C=f|R,S)
t	t	0.9	0.1
t	f	0.8	0.2
f	t	0.7	0.3
f	f	0.1	0.9

Переменные, не имеющие предков описываются безусловными вероятностями, а их потомки на графе - условными

Вычисления в задаче Холмса (продолжение)

- Полные вероятности событий: 1) трава у дома Холмса оказалась влажной, 2) у дома Ватсона наблюдается то же самое.

$$P(C=t)=\sum_{R=\{t,f\},S=\{t,f\}} P(R)\cdot P(S)\cdot P(C=t\,|\,R,S)=$$

$$=0.3\cdot 0.2\cdot 0.9+0.3\cdot 0.8\cdot 0.8+0.7\cdot 0.2\cdot 0.7+0.7\cdot 0.8\cdot 0.1=\underline{0.4}$$

$$P(W=t)=\sum_{R=\{t,f\}} P(R)\cdot P(W=t\,|\,R)=\underline{0.31}$$

Вычисления вероятностей отдельных переменных проводятся путем маргинализации полной вероятности

Вычисления в задаче Холмса (продолжение)

- Если известно, что был дождь, то вероятность наблюдения влажной травы повышается:

$$P\left(C=t \mid R=t\right)=\sum_{S=\{t,f\}} P(S)\cdot P\left(C=t \mid R=t,S\right)=0.82$$

$$P\left(W=t \mid R=t\right)=0.8$$

Условные вероятности вычисляются суммированием по всем возможным значениям переменных, значения которых не определены

Вычисления в задаче Холмса (продолжение)

- Если Холмс выяснил, что трава у дома влажная, то каковы вероятности причин – дождя и поливальной установки? полученные вероятности выше априорных (0.3 и 0.2, соответственно):

$$P(R=t|C=t)=\frac{P(R=t,C=t)}{P(C=t)}=\frac{0.054+0.192}{0.4}=0.615$$

$$P(S=t|C=t)=\frac{P(S=t,C=t)}{P(C=t)}=\frac{0.054+0.098}{0.4}=0.38$$

Условные вероятности причин вычисляются на основе формулы для полной вероятности

Вычисления в задаче Холмса (продолжение)

- Когда Холмс обнаружил, что трава у дома Ватсона также влажная, то вероятности причин изменились!

$$P\left(R=t \middle| C=t, W=t\right)=\frac{P\left(R=t, C=t, W=t\right)}{P\left(C=t, W=t\right)}=$$

$$=\frac{0.8\cdot 0.3\cdot\left(0.18+0.64\right)}{0.8\cdot\left(0.054+0.192\right)+0.1\cdot\left(0.098+0.056\right)}=\frac{0.1968}{0.2122}=0.9274$$

$$P\left(S=t \middle| C=t, W=t\right)=0.2498$$

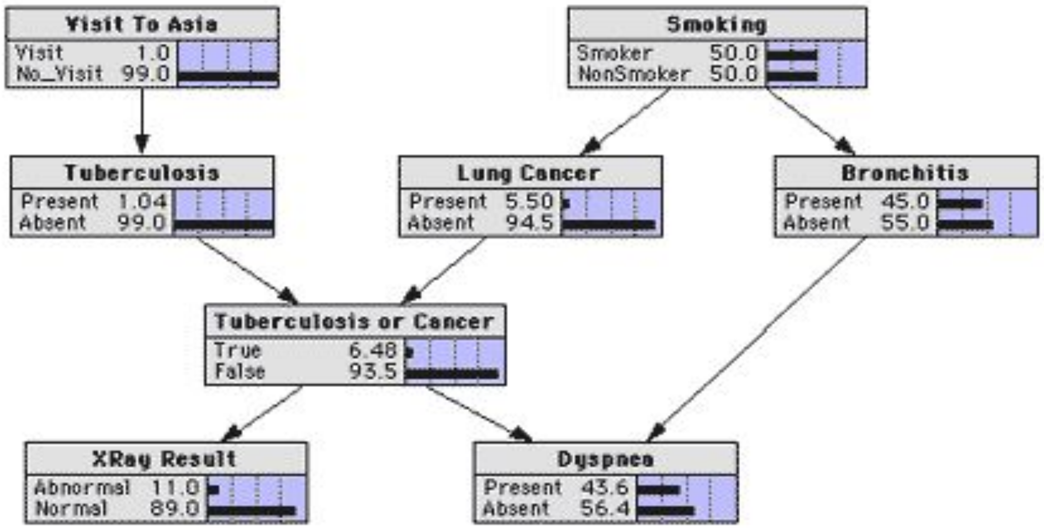
Вероятность не выключенной поливальной установки снизилась почти до первоначального (априорного) значения *автоматически*.

Точные и приближенные вычисления в полномасштабных приложениях

- При росте числа переменных и набора значений отдельных переменных сложность точных вычислений растет комбинаторно.
- На практике широко используются приближенные алгоритмы:
  - Метод Монте-Карло (различные вариации, MCMC, …)
  - Вариационные методы
  - Belief propagation

Некоторые методы позволяют оценить законы распределения, другие обеспечивают только случайные выборки из распределений

Байесова сеть ASIA

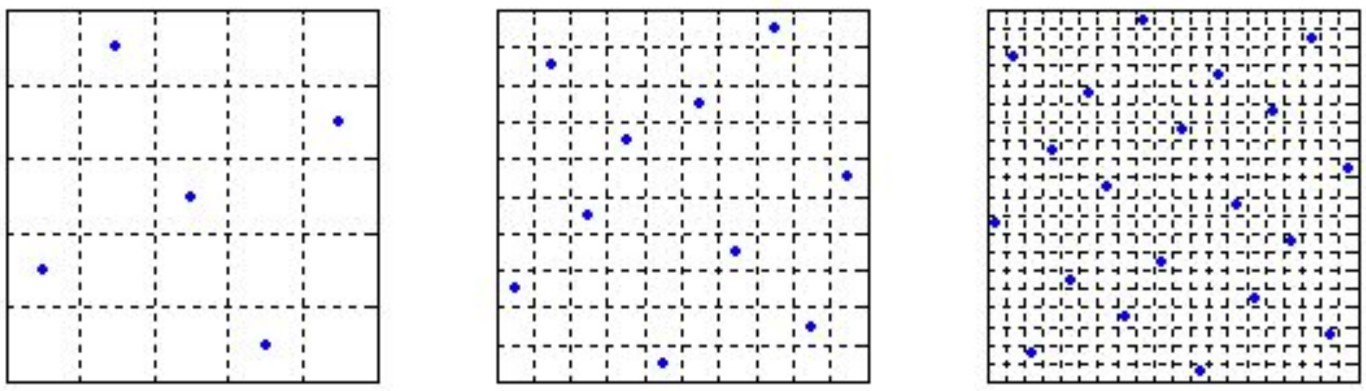


- Учебный пример Байесовой сети в области постановки диагноза (туберкулез, рак легких или бронхит), в зависимости от данных медицинской диагностики (X-ray, удушье, посещение Азии, курение).

Lauritzen, Steffen L. and David J. Spiegelhalter (1988) "Local computations with probabilities on graphical structures and their application to expert systems" in J. Royal Statistics Society B, 50(2), 157-194.

Метод выборок из латинских гиперкубов

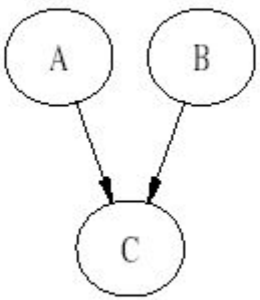
- Выборки из значений многомерных латинских гиперкубов (Latin Hypercube Sampling) – прямое обобщение решения задачи ладей на плоскости.



LHS позволяют моделировать многомерное распределение при фиксированном объеме выборки.

Приближенные вычисления при помощи LHS

Простейшая сеть



Таблицы вероятностей

A	$P(A)$
a_1	0.20
a_2	0.35
a_3	0.45

B	$P(B)$
a_1	0.4
a_2	0.6

$P(C \mid A,B)$	a_1		a_2		a_3	
	b_1	b_2	b_1	b_2	b_1	b_2
c_1	0.01	0.04	0.28	0.06	0.18	0.82
c_2	0.68	0.93	0.52	0.12	0.50	0.10
c_3	0.31	0.03	0.20	0.82	0.32	0.08

Фрагмент гиперкуба

N	A	B	C
1	25	13	74
2	14	91	7
...	...	...	...
39	47	32	56
...	...	...	...
100	69	4	84

- Как при заданных распределения A и B распределена C?
- Для каждой строки гиперкуба используем значение в столбце для получения выборочного значения соответствующей переменной:

$$\begin{cases} a_1, & \mathit{if}\ LHS_{iA} \le P(a_1)\times 100 \\ a_2, & \mathit{if}\ P(a_1)\times 100 \le LHS_{iA} \le \left(P(a_1)+P(a_2)\right)\times 100 \\ a_3, & \mathit{else} \end{cases}$$

- {47,32,56} ->{a2,b1,c2}
- Распределение C дается гистограммой значений для каждой строки

Приближенные вычисления при помощи LHS

- В сложных сетях вычисления проводятся по цепочкам, от предков к потомкам. К розыгрышу очередной переменной можно приступать, когда значения всех ее предков в данном примере выборки уже установлены.
- Сложность вычислений пропорциональна числу переменных и размеру выборки, т.е. определяется размером гиперкуба.
- Результат представляется в форме выборочных гистрограмм распределения интересующих переменных.
- Метод латинского гиперкуба применим и для моделирования непрерывных распределений. В этом случае для розыгрыша безусловных вероятностей используется известный метод обратных функций распределения.

Замечание о субъективных вероятностях и ожиданиях

- Числовые значения вероятностей в Байесовых сетях могут быть как математическими вероятностями, так и субъективными, личностными, оценками ожиданий экспертов по поводу возможности осуществления событий. У разных лиц степень ожидания события может быть разной, это зависит от индивидуального опыта и объема априорной информации.
- Количественная оценка субъективных ожиданий: эксперту предлагается сделать выбор в игре с теоретически известной вероятностью альтернативы. Смена выбора происходит при выравнивании степени ожидания эксперта и теоретической вероятности.
- Субъективные ожидания - единственная альтернатива на практике, когда необходим учет мнения экспертов (например, врачей или социологов).

Синтез Байесовой сети на основе априорной информации

Для построения Байесовой сети необходимо:

- Сформулировать проблему в терминах вероятностей значений целевых переменных;
- Выбрать понятийное пространство задачи, определить переменные, имеющие отношение к целевым переменным, описать возможные значения этих переменных;
- Выбрать на основе опыта и имеющейся информации априорные вероятности значений переменных;
- Описать отношения "причина-следствие" (как косвенные, так и прямые) в виде ориентированных ребер графа, разместив в узлах переменные задачи;
- Для каждого узла графа, имеющего входные ребра указать оценки вероятностей различных значений переменной для комбинаций значений переменных-предков на графе.

Байесово обучение параметров модели по экспериментальным данным

- Если структура связей в сети зафиксирована, то обучение состоит в выборе свободных параметров распределений условных вероятностей.
- *Пример*: Однопараметрическая задача бросания монеты. Параметр θ – вероятность выпадения «орла».
- Пусть в N экспериментах «орел» выпал ровно h раз. Классическая оценка параметра равна выборочной частоте h/N. Это также является классическим прогнозом исхода следующего эксперимента.
- В Байесовом подходе значение параметра само является *случайной* величиной, распределение которой используется при прогнозировании исхода следующего бросания монеты.

Байесово обучение параметров модели по экспериментальным данным

- Функция правдоподобия данных - биномиальное распределение:

$$p\left(D\middle|\theta,\xi\right) \equiv \theta^{h} \cdot (1-\theta)^{t}$$

- Априорная плотность распределения θ – бета-распределение:

$$p\left(\theta\middle|\xi\right) = \beta\left(\theta\middle|\alpha_h,\alpha_t\right) \equiv \frac{\Gamma\left(\alpha_h+\alpha_t\right)}{\Gamma\left(\alpha_h\right)\cdot\Gamma\left(\alpha_t\right)}\theta^{\alpha_h-1}\cdot(1-\theta)^{\alpha_t-1}$$

- Апостериорное распределение параметра дается теоремой Байеса:

$$p\left(\theta\middle|D,\xi\right) = \frac{p\left(D\middle|\theta,\xi\right)\cdot p\left(\theta\middle|\xi\right)}{p\left(D\middle|\xi\right) \equiv \int p\left(D\middle|\theta,\xi\right)\cdot p\left(\theta\middle|\xi\right)d\theta}$$

Прогноз исхода будущего эксперимента по Байесу

- Вероятность выпадения «орла» в будущем эксперименте:

$$p\left(x_{N=1}=H\,\middle|\,D,\xi\right)=\int p\left(x_{N+1}=H\,\middle|\,\theta,\xi\right)\cdot p\left(\theta\,\middle|\,D,\xi\right)d\theta=$$

$$=\int\theta\cdot p\left(\theta\,\middle|\,d,\xi\right)d\theta=\left\langle\theta\right\rangle_{p\left(\theta|D,\xi\right)}$$

- Вычисления для биномиального и бета-распределения:

$$p\left(x_{N+1}=H\,\middle|\,D,\xi\right)=\frac{\alpha_h+h}{\alpha_h+\alpha_t+h+t}$$

- «Физический смысл» результата: использование априорного распределения эквивалентно добавлению “искусственных” (полученных в гипотетических предыдущих экспериментах) отсчетов «орлов» и «решек» в экспериментальную серию.

Обучение параметров Байесовой сети

- Пусть задано множество обучающих примеров *D*, каждый элемент множества – вектор значений для всех переменных Байесовой сети.
- Классическая схема – поиск максимума правдоподобия:

$$L = \frac{1}{N \cdot S} \sum_{j=1}^{N} \sum_{k=1}^{S} \log\left(P\left(x_j \middle| pa\left(x_j \right), D_k \right) \right)$$

- Байесов вариант –априорные вклады в матрицы вероятностей.
- Переменные сети могут могут обучаться на отдельных наборах примеров, учитывающих значения только тех переменных, которые влияют на данную.
- Условные вероятности в Баейсовой сети могут быть представлены нейронной сетью или другими удачными аппроксимациями плотности вероятности.

Представление распределений в Байесовых сетях вероятностными деревьями

- Пусть имеется одна зависимая переменная $p\left(y\,|\,\vec{x}\right)$ и задано множество обучающих примеров $D:\{\vec{X};Y\}$.
- Выберем дискретизацию переменной y так, что в каждый отрезок попадает одинаковое число наблюдений.

$$y_{\min} < y_1 < \ldots < y_k < \ldots < y_{\max}$$

- Априорная заселенность всех интервалов одинакова, что соответствует максимуму энтропии:

$$S_0 = -N_0 \sum_k p_k \log p_k$$

Построение энтропийного дерева

- Вся совокупность данных образует *корень дерева*, на котором (максимальное) значение энтропии отвечает полному отсутствию информации о возможном значении зависимой переменной.
- Каждому узлу дерева (включая корень) приписывается решающее правило, разделяющее множество данных на два подмножества.
- Правило выбирается из условия максимального уменьшения суммарной энтропии подмножеств:

$$S_1 = S_1^{'} + S_1^{''} < S_0$$

- Каждое правило является *максимально информативным* – остаточная энтропия после его применения *минимальна*.

Правила в узлах энтропийного дерева

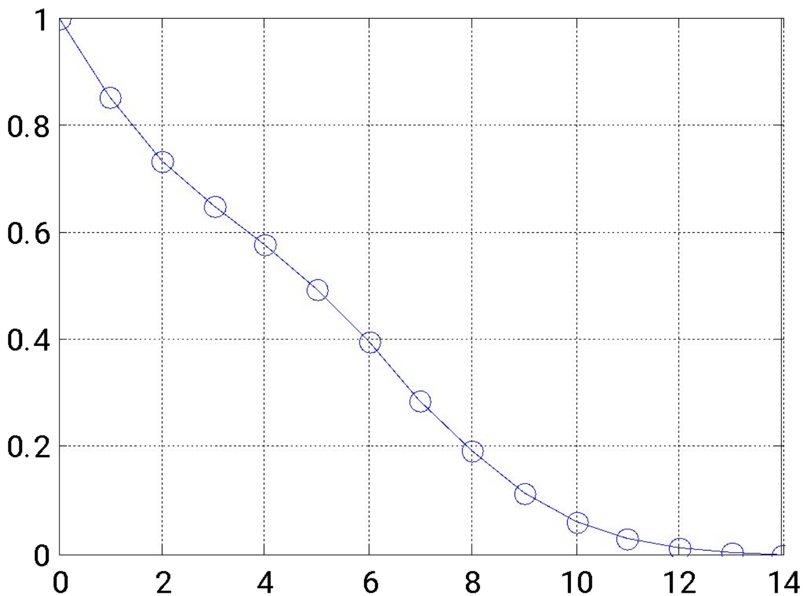


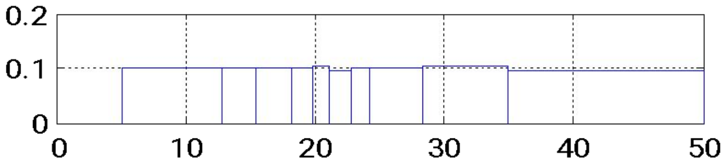
Рис. Зависимость энтропии (в долях к исходному значению) от уровня иерархии дерева

- Простейший класс правил – пороговый выбор по значению одного из аргументов. Наилучший аргумент определяется путем решения серии задач одномерной оптимизации.
- Иерархический процесс далее продолжается для подмножеств каждого узла. Процесс формально завершается по достижении нулевой энтропии для каждого узла самого нижнего уровня.

Структура энтропийного дерева

- В итоге, каждому узлу полученного дерева приписывается:
- Эмпирическая оценка плотности условного распределения дискретизованной зависимой переменной (при условии отнесения примера к данному узлу);
- Оценка выборочной энтропии распределения в этом узле;
- Решающее правило, позволяющее выбрать дочернюю ветвь с дальнейшим уменьшением энтропии условного распределения.

Плотность распределения в иерархиях энтропийного дерева



- Представлена плотность вероятности распределения зависимой переменной при условии отнесения примера к данному узлу дерева
- При обучении может использоваться регуляризирующий критерий останова по предельной сложности дерева

Свойства энтропийных деревьев

- При дроблении множества данных до нуля энтропии полученное дерево является, очевидно, переобученным, т.к. в нем полностью запомнен весь шум, содержавшийся в данных. На практике для оценок обобщающей способности можно воспользоваться методикой кросс-валидации на основе бутстрэп-выборок.
- Метод всегда сходится, по крайней мере, за (N-1) шагов.
- Вычислительная сложность метода невысока и ограничивается решением задач одномерной оптимизации на отрезке.

Информационная значимость факторов

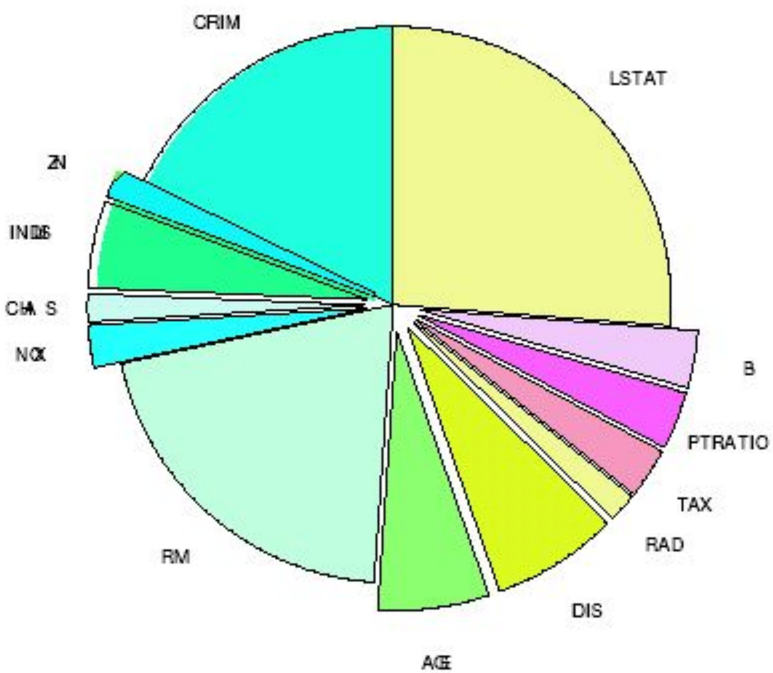


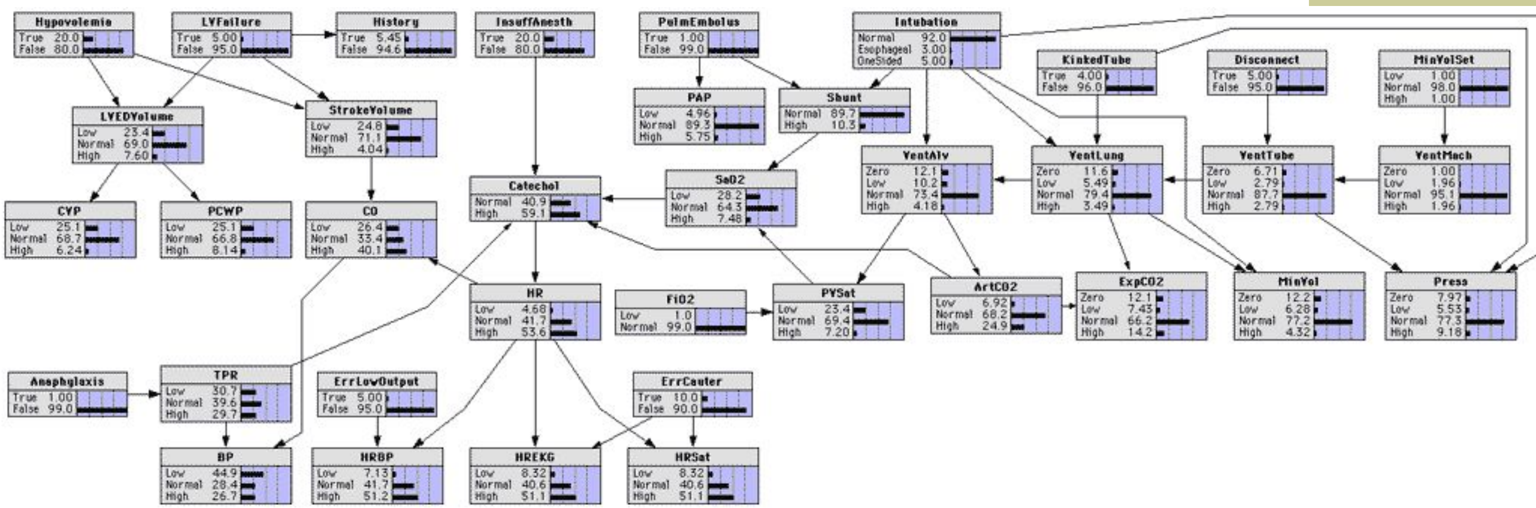
Рис. Относительная значимость факторов в задаче Boston Housing

- Информационный вклад различных входных переменных в снижение энтропии весьма неоднороден.
- Все факторы (независимые переменные) можно упорядочить по степени уменьшения суммарной энтропии зависимой переменной, возникшей при использовании фактора в решающих правилах. Переменные, не использованные ни в одном из правил малозначимы.

Примеры приложений Байесовых сетей

- Естественной областью использования Байсовых сетей являются экспертные системы, которые нуждаются в средствах оперирования с вероятностями.
- **Медицина**
  - Система PathFinder (Heckerman, 1990) разработана для диагностики заболеваний лимфатических узлов. PathFinder включает 60 различных вариантов диагноза и 130 переменных, значения которых могут наблюдаться при изучении клинических случаев. Система смогла приблизиться к уровню экспертов, и ее версия PathFinder-4 получила коммерческое распространение.
  - Множество других разработок (Child, MUNIN, Painulim, SWAN и пр.) успешно применяются в различных медицинских приложениях [Jensen96].

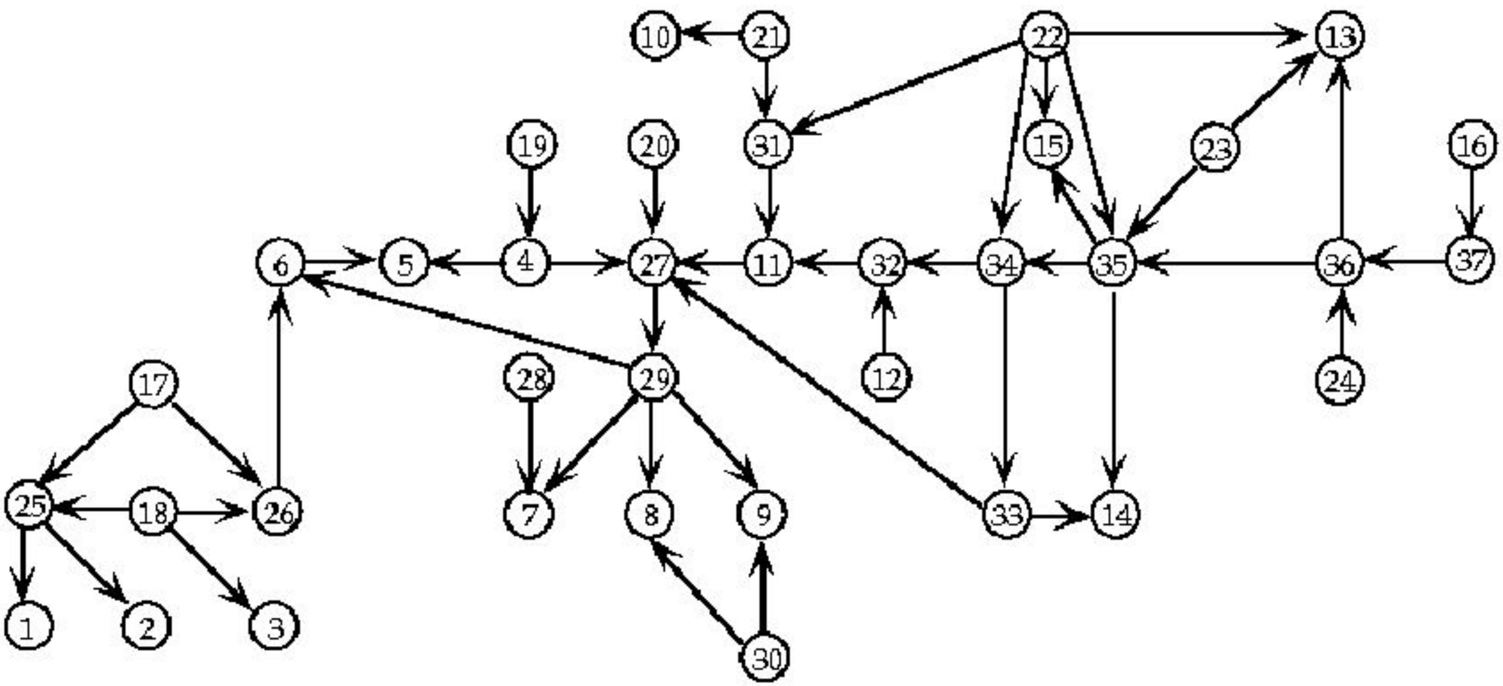
Байесова сеть ALARM в области медицинской диагностики



- Диагностическая система ALARM (A Logical Alarm Reduction Mechanism) в области диагностики и мониторинга состояния пациента.

BeinlichSCC89: Beinlich, Ingo, H. J. Suermondt, R. M. Chavez, and G. F. Cooper (1989) "The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks" in Proc. of the Second European Conf. on Artificial Intelligence in Medicine (London, Aug.), 38, 247-256. Also Tech. Report KSL-88-84, Knowledge Systems Laboratory, Medical Computer Science, Stanford Univ., CA.

Логическая архитектура Байесовой сети в полномасштабном приложении ALARM (медицина)



Описание переменных сети ALARM

1 – central venous pressure
2 – pulmonary capillary wedge pressure
3 – history of left ventricular failure
4 – total peripheral resistance
5 – blood pressure
6 – cardiac output
7 – heart rate obtained from blood pressure monitor
8 – heart rate obtained from electrocardiogram
9 – heart rate obtained from oximeter
10 – pulmonary artery pressure
11 – arterial-blood oxygen saturation
12 – fraction of oxygen in inspired gas
13 – ventilation pressure
14 – carbon-dioxide content of expired gas
15 – minute volume, measured
16 – minute volume, calculated
17 – hypovolemia
18 – left-ventricular failure
19 – anaphylaxisr
20 – insufficient anesthesia or analgesia
21 – pulmonary embolus
22 – intubation status
23 – kinked ventilation tube
24 – disconnected ventilation tube
25 – left-ventricular end-diastolic volume
26 – stroke volume
27 – catecholamine level
28 – error in heart rate reading due to low cardiac output
29 – true heart rate
30 – error in heart rate reading due to electrocautery device
31 – shunt
32 – pulmonary-artery oxygen saturation
33 – arterial carbon-dioxide content
34. – alveolar ventilation
35 – pulmonary ventilation
36 – ventilation measured at endotracheal tube
37 – minute ventilation measured at the ventilator

Примеры приложений Байесовых сетей

- **Космические и военные применения**
  - Система поддержки принятия решений Vista (Eric Horvitz) применяется в Центре управления полетами NASA (NASA Mission Control Center) в Хьюстоне. Система анализирует телеметрические данные и в реальном времени идентифицирует, какую информацию нужно выделить на диагностических дисплеях.
  - В исследовательской лаборатории МО Австралии Байесовы сетиприменяются в тактических задачах исследования операций. Учебная модель «Operation Dardanelles» охраны территориальной зоны с моря включает в себя различные тактические сценарии поведения конфликтующих сторон, данные о передвижении судов, данные разведнаблюдений, и другие переменные. Последовательное поступление информации о действиях противников позволяет синхронно прогнозировать вероятности различных действий в течение конфликта.

Anti-Air Treat Identification Problem

Variable	Description	# States	Parents
T	Threat ID	7	{Ø}
R	Range	7	{Ø}
E	Emitter	7	{T}
TM	Threat Mode	7	{T, R}
G	Guidance	2	{T}
V	Visibility	3	{Ø}
ML	Missile Launch Indicator	2	{TM}
EO	Electro-Optical Sensor	3	{TM, G}
RWR	Radar Warning Receiver	2	{TM, G}
TE	Threat Effectiveness	2	{G, V}

Переменные задачи

Связи между переменными

$R \perp T$

$E \perp R | T$

$TM \perp E \mid \{T, R\}$

$G \perp \{R, E, TM\} \mid T$

$V \perp \{T, R, E, TM, G\}$

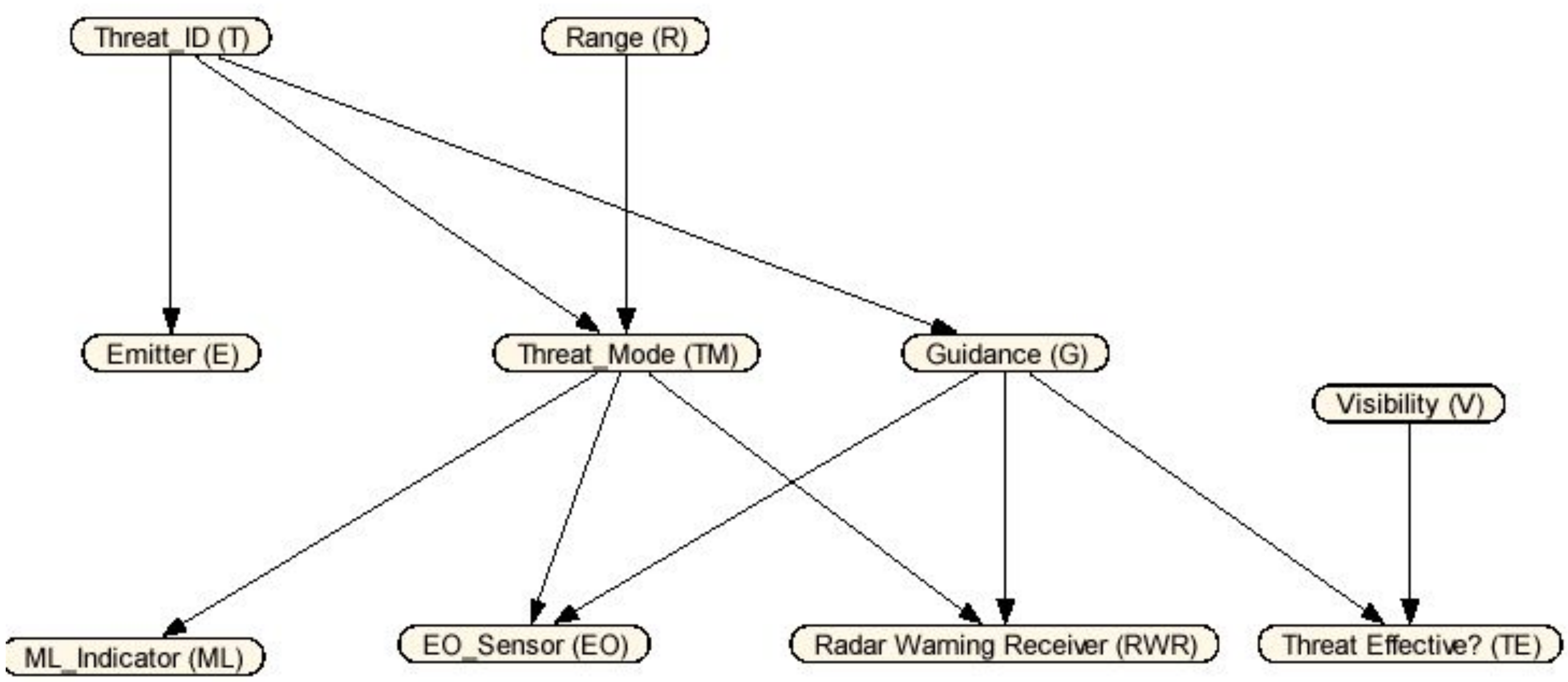
$ML \perp \{T, R, E, G, V\} \mid TM$

$EO \perp \{T, R, E, V, ML\} \mid \{TM, G\}$

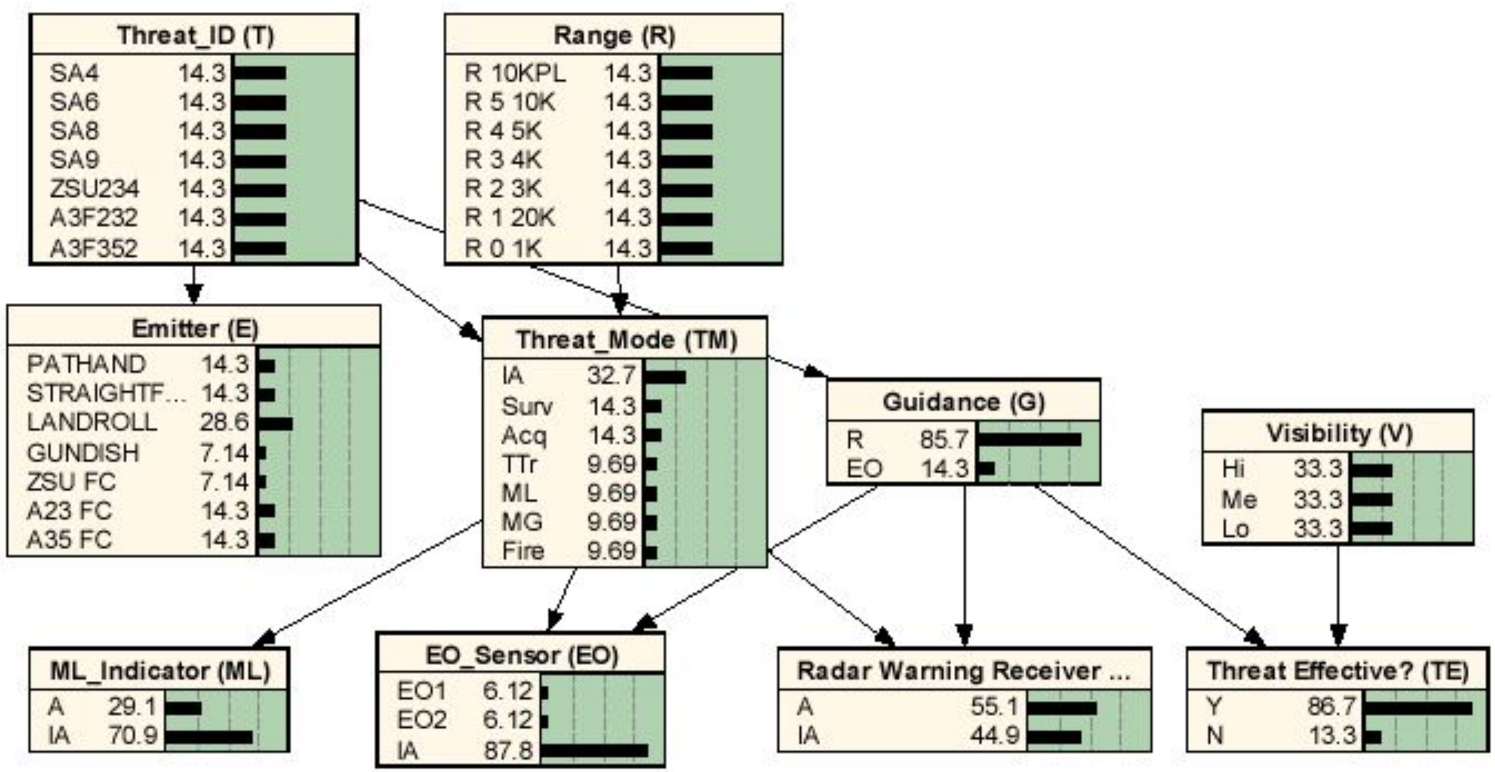
$RWR \perp \{T, R, E, V\} \mid \{TM, G\}$

$TE \perp \{T, R, E, TM, ML, EO, RWR\} \mid \{G, V\}$

Anti-Air Treat Identification Problem



Anti-Air Treat Identification Problem



Примеры приложений Байесовых сетей

- **Компьютеры и системное программное обеспечение**
  - Microsoft: управление интерфейсными агентами в системе Office, в диагностике проблем работы принтеров и других справочных и wizard-подсистемах.
- **Обработка изображений и видео**
  - Восстановление 3D сцен из динамической 2D информации, синтез статических изображений высокой четкости из видеосигнала.
- **Финансы и экономика**
  - Оценка риска и прогноз доходности портфелей финансовых инструментов. Основные достоинства Байесовых сетей: возможность совместного учета количественных и качественных рыночных показателей, динамическое поступление новой информации, а также явные зависимости между существенными факторами, влияющими на финансовые показатели, наглядное игровое моделирование.

Компьтерный пакет Netica

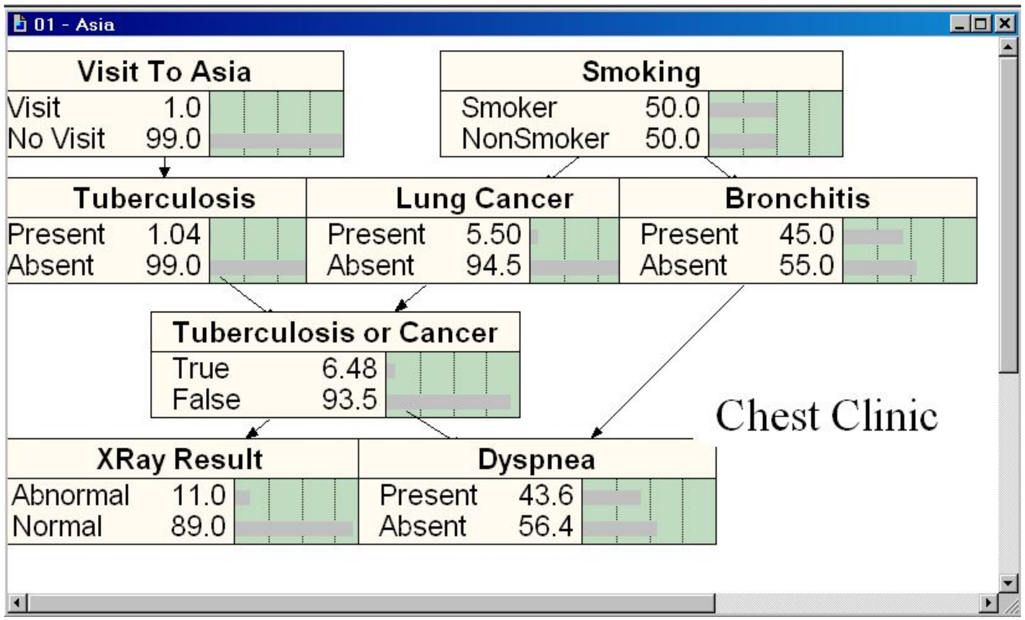


Рис. Пример интерфейса системы NETICA http://www.norsys.com/

- Netica - мощная, удобная в работе программа для работы с графовыми вероятностными моделями. Коммерчески доступна с 1995 г. В настоящее время Netica является одним из наиболее широко используемых инструментов для разработки Байсовых сетей.

Ресурсы Интернет по Байесовым сетям

- http://www.auai.org/ - Ассоциация Анализа Неопределенности в Искусственном Интеллекте (Association for Uncertainty in Artificial Intelligence - AUAI)
- http://www.norsys.com/ - Norsys Software Corp (NETICA)
- http://www.kic.com/ - Компания Knowledge Industries
- http://www.data-digest.com/ - Data Digest Corporation
- http://www.hugin.com/ - Компания HUGIN Expert
- http://www.bayesware.com/ - Компания BayesWare, Ltd
- http://stat.rutgers.edu/~madigan/bayes_people.html - Персональные страницы специалистов по Байесовым методам

Итоги

- Байесовы вероятностные методы обучения машин являются существенным шагом вперед, в сравнении с популярными моделями "черных ящиков". Они дают понятное объяснение своих выводов, допускают логическую интерпретацию и модификацию структуры отношений между переменными задачи, а также позволяют в явной форме учесть априорный опыт экспертов.
- Благодаря удачному представлению в виде графов, Байесовы сети весьма удобны в пользовательских приложениях.
- Байесовы сети базируются на фундаментальных положениях и результатах теории вероятностей, разрабатываемых в течение нескольких сотен лет, что и лежит в основе их успеха в практической плоскости. Редукция совместного распределения вероятностей в виде произведения условных вероятностей, зависящих от малого числа переменных, позволяет избежать “комбинаторных взрывов” при моделировании.

Перспективы

- Байесова методология, в действительности, шире, чем семейство способов оперирования с условными вероятностями в ориентированных графах. Она включает в себя также модели с симметричными связями (случайные поля и решетки), модели динамических процессов (Марковские цепи), а также широкий класс моделей со скрытыми переменными, позволяющих решать вероятностные задачи классификации, распознавания образов и прогнозирования.
- Новые области применений:
  - Динамические процессы и динамическое программирование
  - Оптимальное управление стохастическими системами
  - Принятие решений в автономных интеллектуальных системах