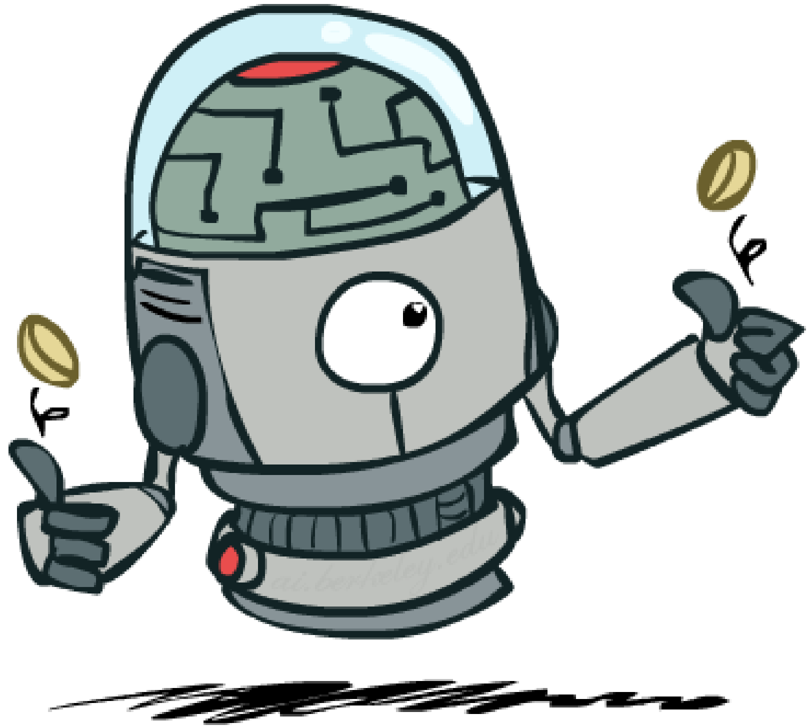
Севастопольский государственный университет
Институт информационных технологий

Методы и системы искусственного интеллекта

Бондарев Владимир Николаевич

Байесовские сети: представление

Независимость



Независимость

- Две переменные (абсолютно) независимы, если:

$$\forall x, y : P(x, y) = P(x)P(y)$$

  - Это говорит о том, что их совместное распределение **факторизуется** в виде произведения двух более простых распределений.
  - Другая форма независимости:

$$\forall x, y : P(x|y) = P(x)$$

  - Независимость *X* и *Y* обозначают в виде: $X \perp\!\!\!\perp Y$

- Если множество всех переменных может быть разделено на независимые подмножества, то полное совместное распределение может быть **факторизовано** в виде произведения отдельных совместных распределений, заданных на этих подмножествах.
- *Независимость - это упрощающее допущение при моделировании:*
  - Что можно предположить в отношении независимости следующих переменных: {Погода, Траффик, Кариес, Зубная боль}?

Пример: Независимость?

Для какого из совместных распределений, переменные T и W независимы?

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T,W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P_2(T,W)$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

✓ _

$P(W)$

W	P
sun	0.6
rain	0.4

Пример: Независимость

- *n* независимых подбрасываний монеты :

$P(X_1)$

H	0.5
T	0.5

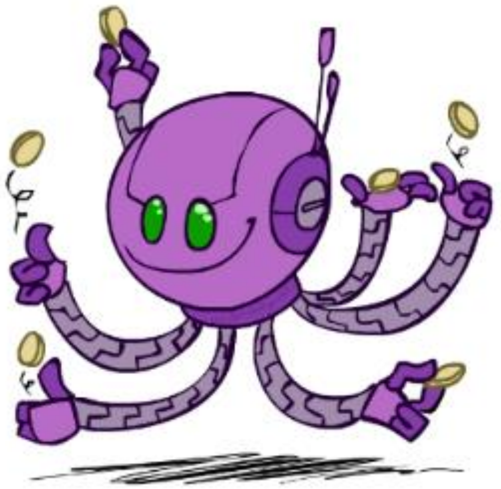
$P(X_2)$

H	0.5
T	0.5

$\cdots$

$P(X_n)$

H	0.5
T	0.5



Совместное распределение результатов *n* независимых подбрасываний монеты

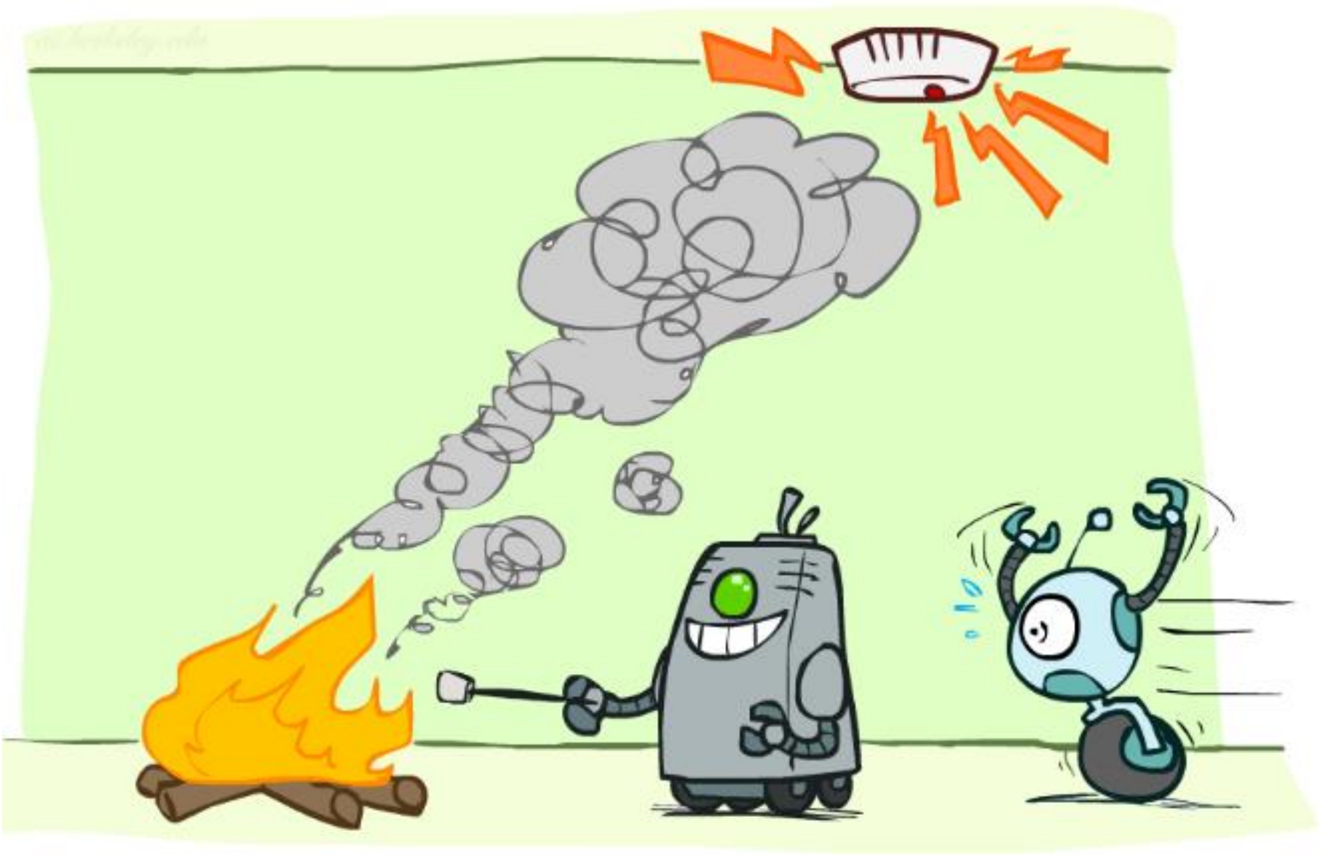
$P(X_1, X_2, \ldots X_n)$

2^n

Совместное распределение результатов *n* независимых бросков монеты $P(X_1, X_2,\ldots,X_n)$ может быть представлено как произведение *n* распределений $P(X_i)$ с одной переменной. В этом случае пространственная сложность совместного распределения будет пропорциональна 2*n.

На практике подобная независимость встречается редко. И тут на помощь приходит условная независимость.

Условная независимость



Условная независимость

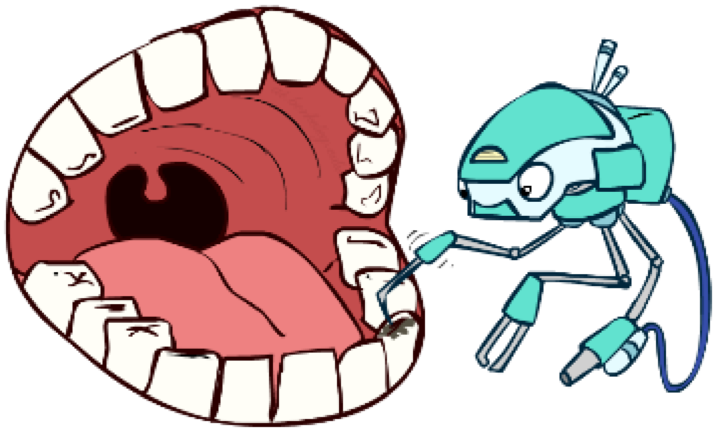
Рассмотрим пример:

- P(Toothache, Cavity, Catch)

- Если у больного есть «дупло», вероятность того, что зонд зацепится за зуб, не зависит от того, болит ли зуб (+toothache):
  - P(+catch | +toothache, +cavity) = P(+catch | +cavity)

- Такая же независимость от боли сохраняется, если нет «дупла»:
  - P(+catch | +toothache, -cavity) = P(+catch| -cavity)

- Т.о., Catch ***условно не зависит*** от зубной боли:
  - P(Catch | Toothache, Cavity) = P(Catch | Cavity)
- Условная независимость переменных Toothache и Catch :
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - **P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)**



Здесь мы используем менее радикальное условие, чем полная независимость. В итоге мы принимаем предположение условной независимости

Toothache _||_ Catch | Cavity

Условная независимость

- *Условная независимость* - это основная форма знаний в условиях неопределенности.

- Общее определение условной независимости двух переменных X и Y, если дана третья переменная Z:

X условно не зависит от Y при заданном Z, $X \perp\!\!\!\perp Y | Z$
если и только если:

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

или, эквивалентно, если и только если

$$\forall x, y, z : P(x|z, y) = P(x|z)$$

$$P(x|z, y) = \frac{P(x, z, y)}{P(z, y)}$$

$$= \frac{P(x, y|z)P(z)}{P(y|z)P(z)}$$

$$= \frac{P(x|z)P(y|z)P(z)}{P(y|z)P(z)}$$

Условная независимость

- Что можно сказать о независимости в следующих переменных :
  - Траффик
  - Зонтик (umbrella)
  - Дождь (raining)

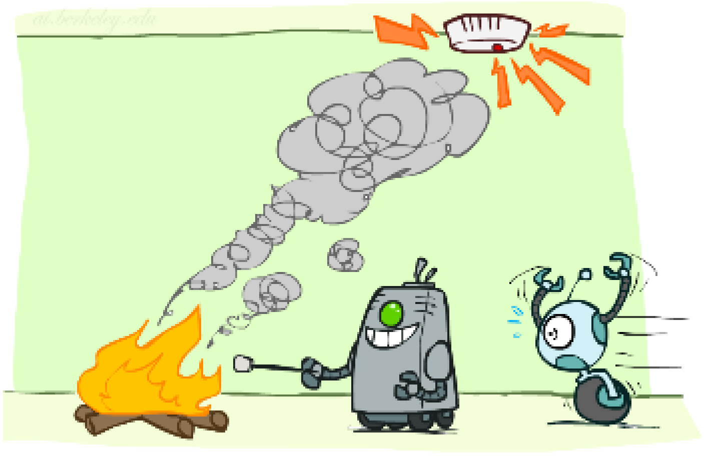
T_||_U | R , т.е. T и U независимы , если идет дождь.

Если идет дождь, то закрытие зонта не уберет пробки. Значит T и U условно независимы.

Условная независимость

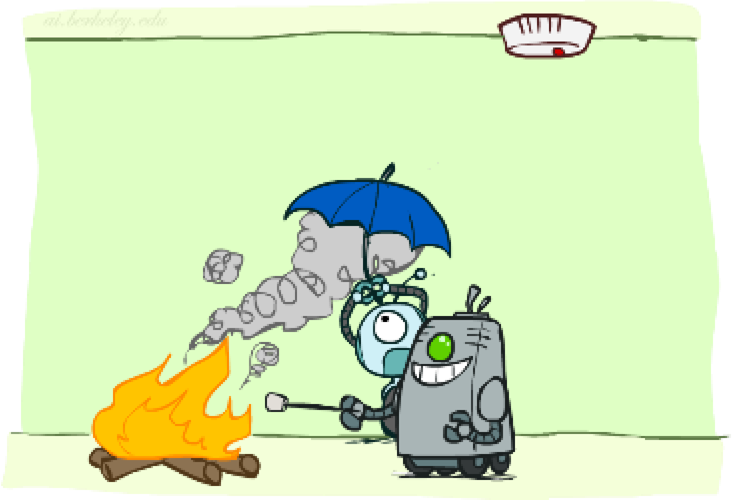
- Что можно сказать об этих переменных :
  - Огонь
  - Дым
  - Тревога

  Рассмотрите рисунки.



Тревогу создает Дым. Поэтому Огонь и Тревога – условно независимы

Огонь_||_Тревога | Дым



При наличии **Дыма** не важно по какой причине он там появился.

Условная независимость и цепочное правило

- Цепочное правило: $P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$

- Тривиальная декомпозиция:

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic})$$

- В случае условной независимости T_||_U | R :

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$

Утверждения с описанием свойств абсолютной независимости позволяют выполнять декомпозицию полного совместного распределения на гораздо более мелкие распределения.
Аналогичную декомпозицию позволяют выполнять утверждения об условной независимости.
Таким образом, первоначальная крупная таблица **P(Traffic, Rain, Umbrella)** декомпонована на три меньшие таблицы.

Условная независимость и цепочное правило

- Цепочное правило: $P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$
- Тривиальная декомпозиция:

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic})$$

- В случае условной независимости T_||_U | R :

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$

Разработка методов декомпозиции крупных вероятностных областей определения на слабо связанные подмножества с помощью свойства условной независимости стало одним из наиболее важных достижений в новейшей истории искусственного интеллекта.

Утверждения с описанием свойств абсолютной независимости позволяют выполнять декомпозицию полного совместного распределения на гораздо более мелкие распределения.
Аналогичную декомпозицию позволяют выполнять утверждения об условной независимости.
Таким образом, первоначальная крупная таблица **P(Traffic, Rain, Umbrella)** декомпонована на три меньшие таблицы.

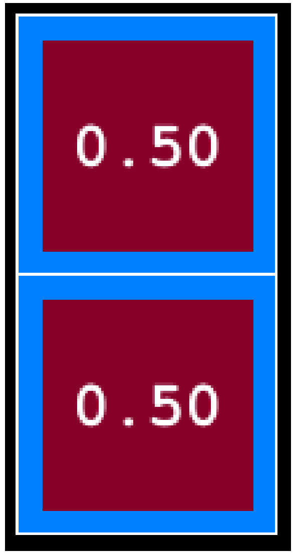
Цепочное правило при поиске призраков

- Призрак может находиться в верхней или нижней ячейке. Сообщение сенсора зависит только от локации призрака.

- Это означает, что сообщения сенсора условно независимы при заданной локации призрака

- T: Квадрат сверху красный (top)
  B: Квадрат снизу красный (bottom)
  G: Призрак вверху

- Дано:
  P(+g) = 0.5
  P(-g) = 0.5
  P(+t | +g) = 0.8
  P(+t | -g) = 0.4
  P(+b | +g) = 0.4
  P(+b | -g) = 0.8



Нам даны частные распределения.
Мы хотим получить совместное распределение
P(T,B,G)=P(G)*P(T|G)*P(B|T,G). Но последнее нам неизвестно.

Если учесть B_||_T|G, то
P(T,B,G) = P(G) P(T|G) P(B|G)

T	B	G	P(T,B,G)
+t	+b	+g	0.16
+t	+b	-g	0.16
+t	-b	+g	0.24
+t	-b	-g	0.04
-t	+b	+g	0.04
-t	+b	-g	0.24
-t	-b	+g	0.06
-t	-b	-g	0.06

Байесовские сети

Байесовские сети - графические модели, которые помогают эффективно выражать допущения условной независимости

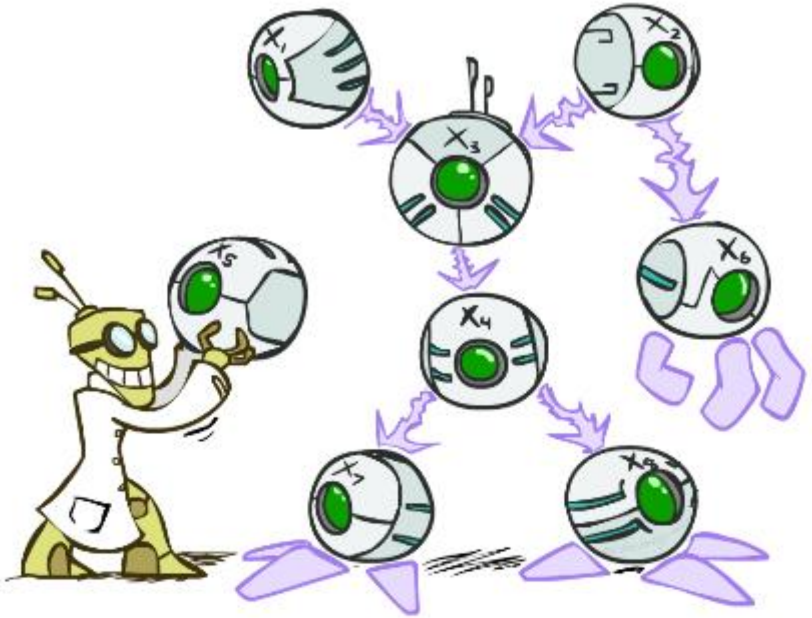
Такое название применяется наиболее часто, но для обозначения этой структуры данных предусмотрено также много других названий:

- **сеть доверия** (belief network);
- **вероятностная сеть** (probabilistic network);
- **причинная сеть** (causal network) ;
- **схема знаний** (knowledge map).

В статистике термином **графическая модель** (graphical model) обозначается немного более широкий класс структур данных, который включает байесовские сети. Еще одно расширение байесовских сетей, называемое **сетью принятия решений** (decision network), или **диаграммой влияния** (influence diagram),

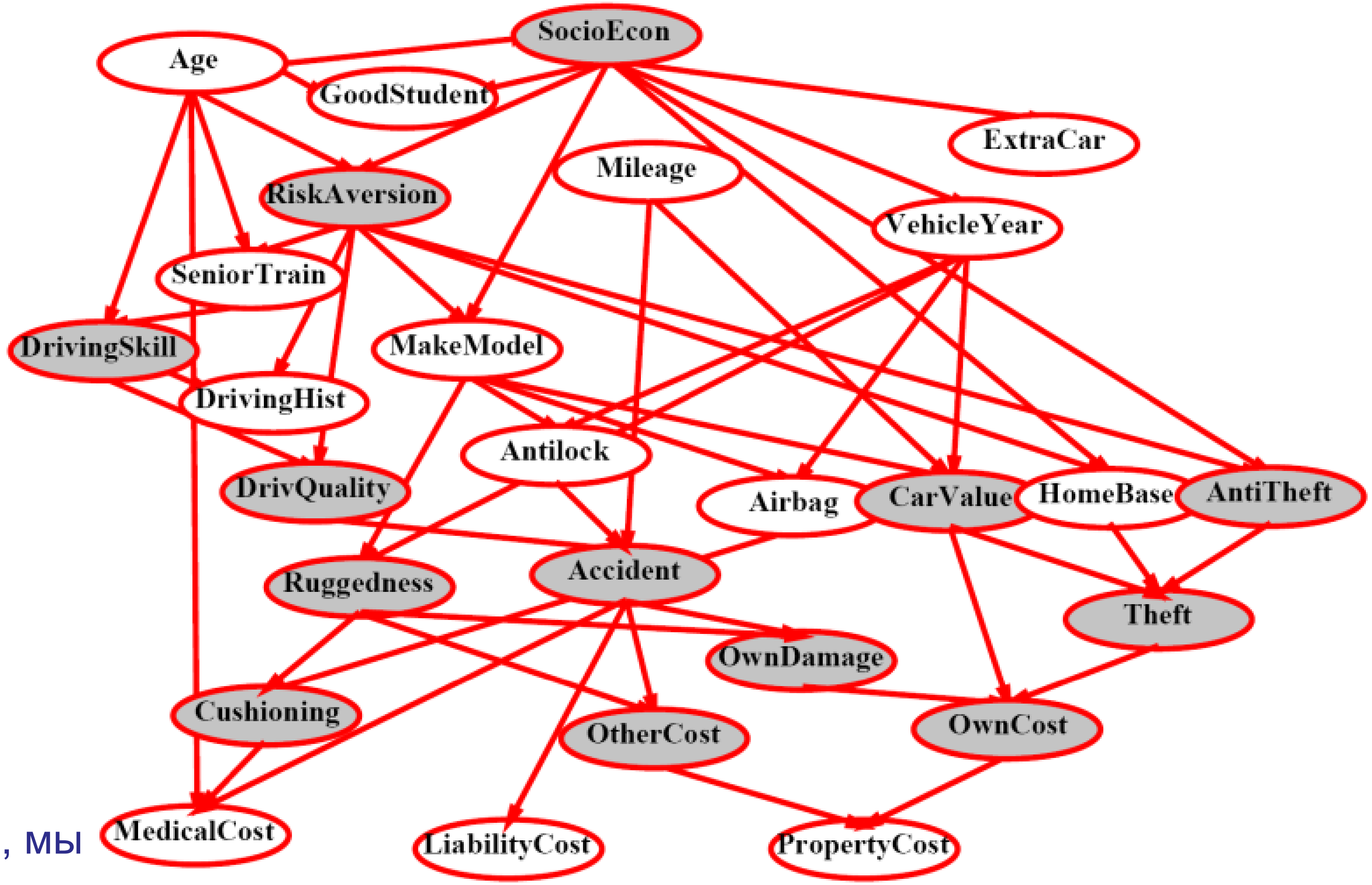
Байесовские сети

- Две проблемы, возникающие при использовании полных совместных таблиц распределений в качестве вероятностных моделей :
  - Даже если имеется небольшое число переменных, совместное распределение слишком велико, чтобы его можно было представить в явном виде;
  - Трудно исследовать (оценить) что-либо эмпирическим путем, если рассматриваются несколько переменных одновременно.



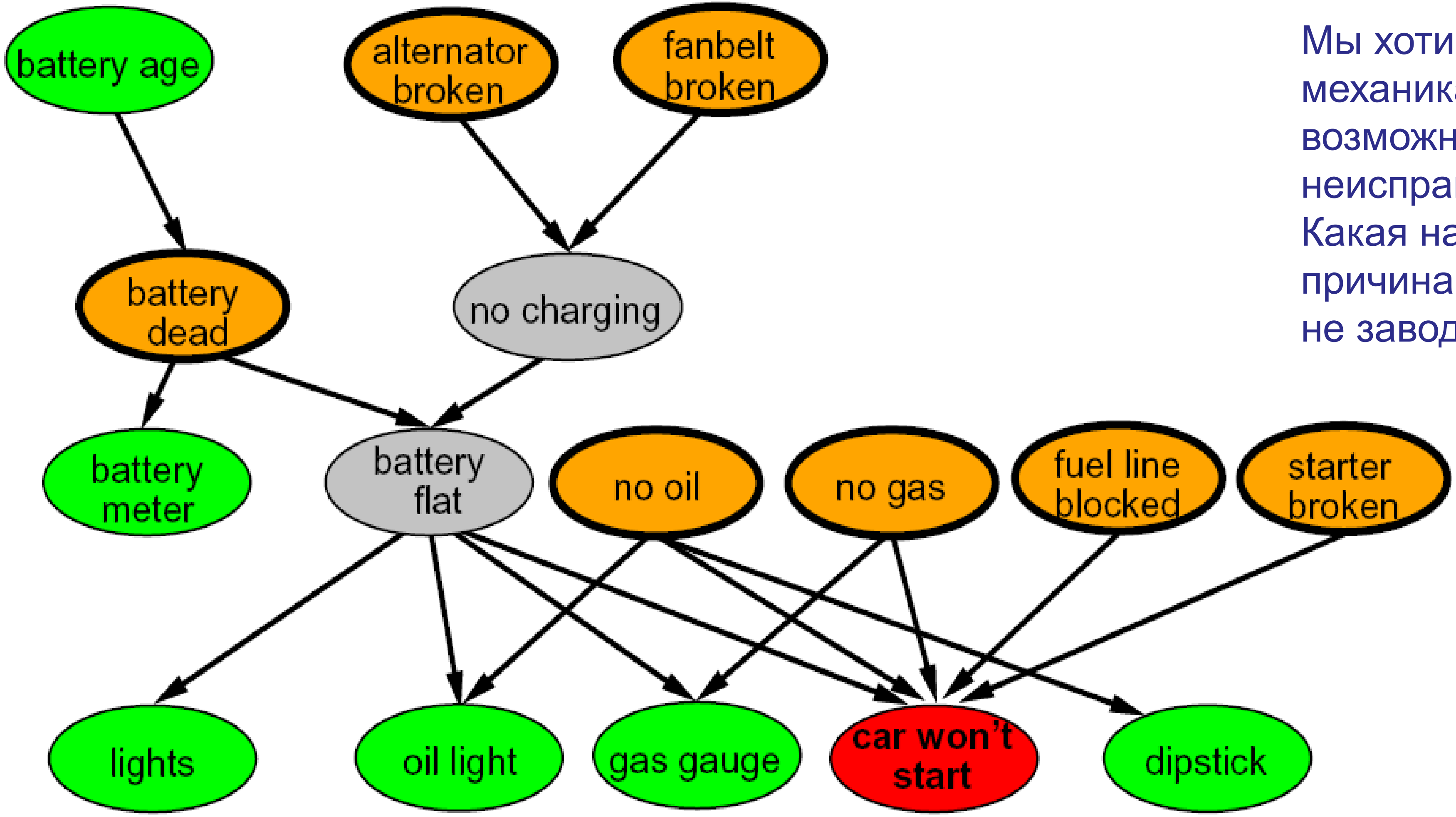
- Сети Байеса: метод описания сложных совместных распределений (моделей) с использованием простых локальных распределений (условных вероятностей):
  - Более правильно называть графическими моделями;
  - Описывают локальное взаимодействие переменных;
  - Локальные взаимодействия объединяются в цепочку, чтобы сформировать глобальные непрямые взаимодействия.

Пример сети Байеса: Страхование



30 переменных с 10 значениями, т.е. размер распределения 10^30!!! Какова вероятность дтп (Accident)? Благодаря локальным связям и условной независимости, мы можем её вычислить.

Пример сети Байеса : Автомобиль

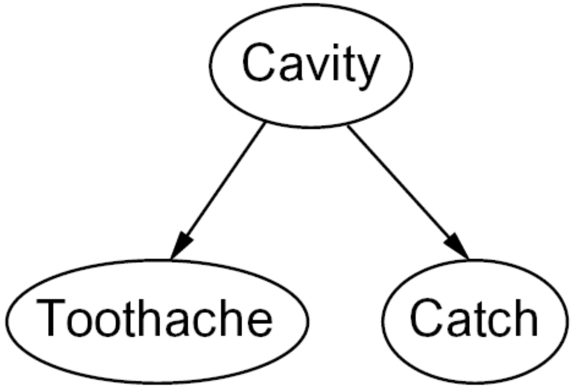


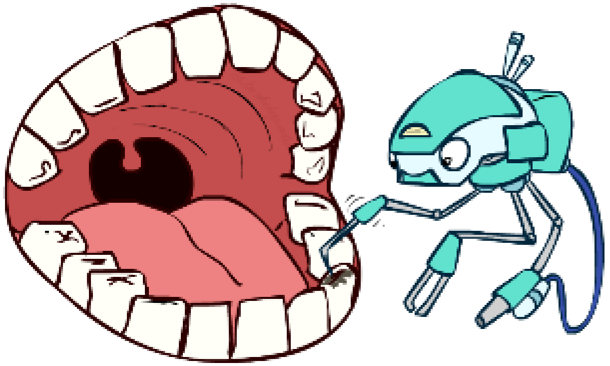
Мы хотим создать робота механика. Зеленым указаны возможные признаки неисправностей (свидельства), Какая наиболее вероятная причина того, что автомобиль не заводится (car won't start)?

Нотации графических моделей

- Байесовская сеть — это ориентированный граф, в котором каждая вершина снабжена количественной вероятностной информацией.

- Вершины - случайные переменные
  - Переменные могут быть дискретными или непрерывными.
- Ребра: обозначают взаимодействия
  - Подобно ограничениям CSP ;
  - Обозначают прямое влияние одной переменной на другую.



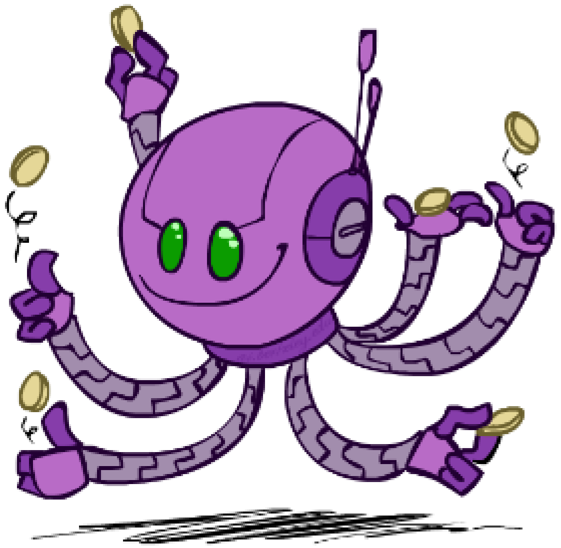


- Топология сети: (множество вершин и ребер) показывает отношения, определяющие условную независимость.

Toothache _||_ Catch | Cavity

Пример: подбрасывание монеты

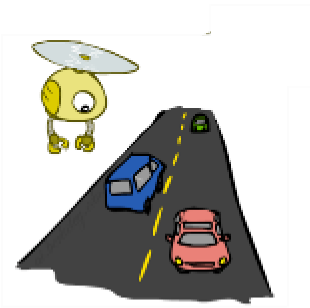
- N независимых подбрасываний монеты

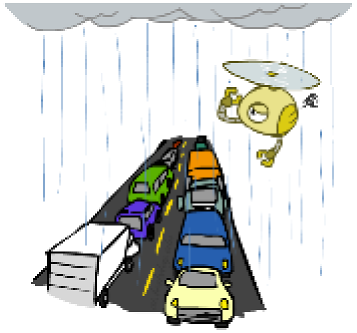


- Нет взаимодействий между переменными: абсолютная независимость

Пример: Траффик I

- Переменные:
  - R: Дождь
  - T: Пробка





- Модель 1: независимость

$$\begin{tikzcd} R \\ T \end{tikzcd}$$

- Модель 2: дождь приводит к пробкам

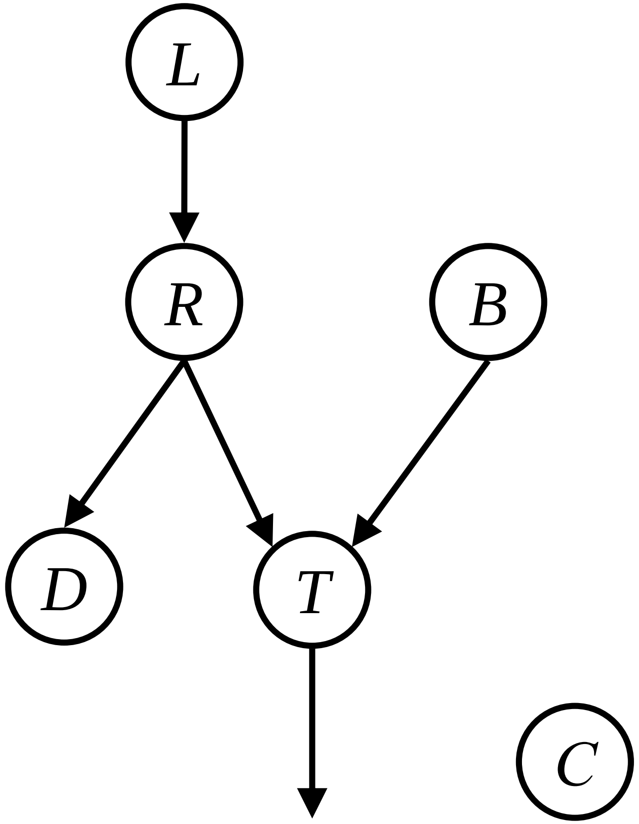
$$\begin{tikzcd} R \arrow[d] \\ T \end{tikzcd}$$

Модель 2: Поведение T – функция от R.
Вторая модель лучше, потому что позволяет делать выводы:
по R предсказывает T.

- Почему агенту лучше использовать модель 2?

Пример: Трафик II

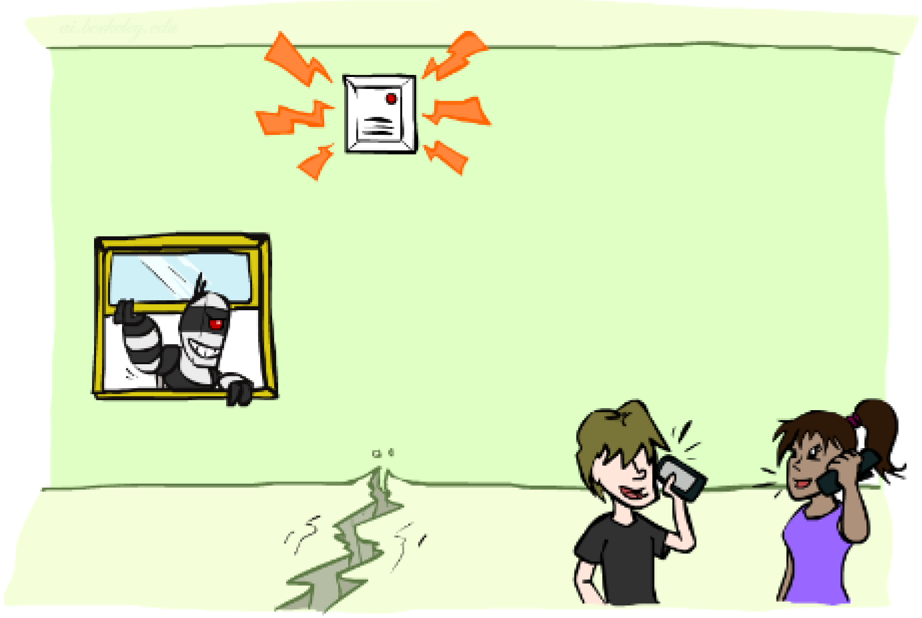
- Переменные
  - T: Пробка
  - R: Дождь
  - L: Низкое давление
  - D: Капли по крыше
  - B: Игра в футбол
  - C: Кариес



Изобразим сеть Байеса для ситуации, представленной на рисунке

Пример: Сеть тревоги

Житель пригорода установил систему тревожной сигнализации для обнаружения взлома. Она иногда также реагирует на небольшие землетрясения. Его соседи, Джон и Мэри, обещали звонить ему на работу, услышав тревожный сигнал. Джон иногда путает тревожный сигнал с телефонным звонком в доме соседа и в этих случаях также звонит. Мэри любит слушать музыку и поэтому иногда вообще пропускает тревожный сигнал. Получив факты о том, кто из соседей звонил или не звонил, необходимо оценить вероятность взлома.

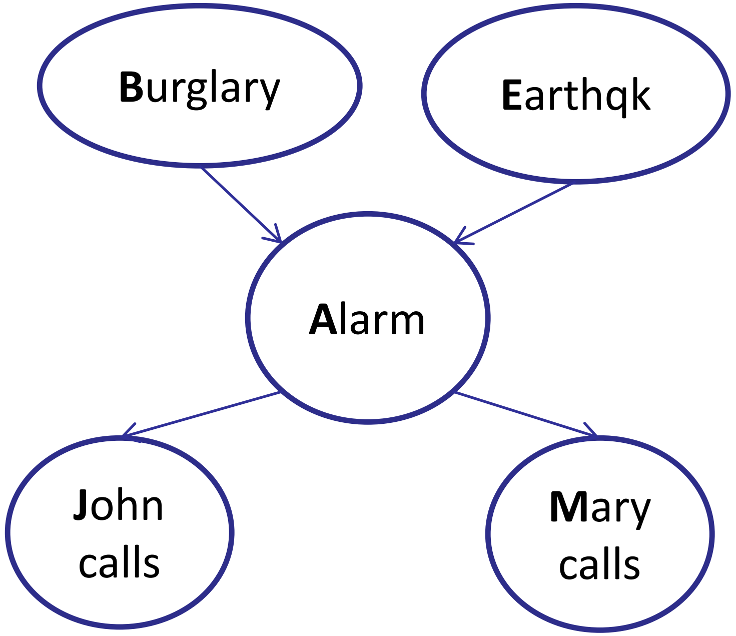


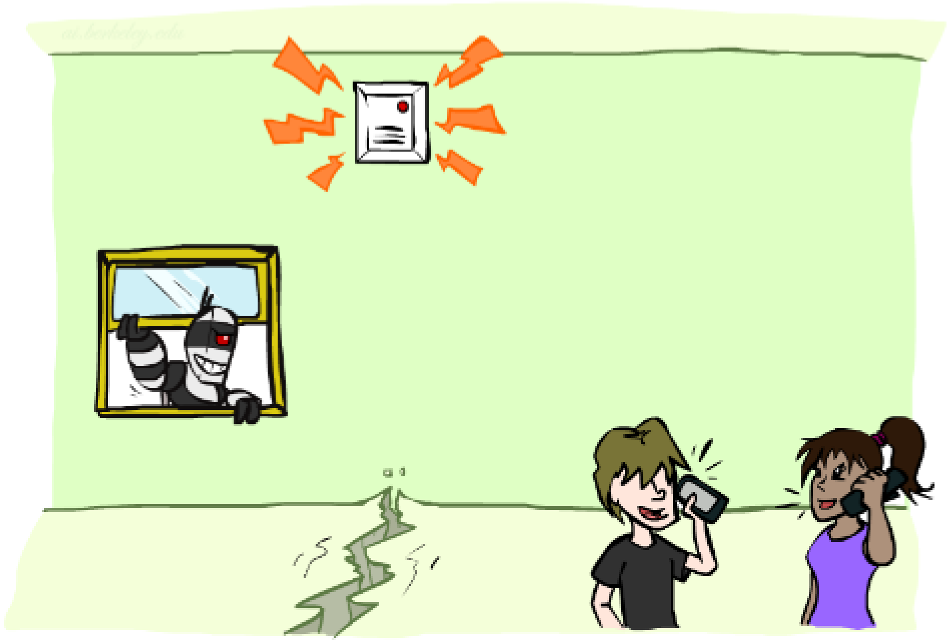
- Переменные
  - B: Ограбление
  - A: Тревога
  - M: Звонок Мэри
  - J: Звонок Джона
  - E: Землетрясение

Этим примером мы обязаны **Джуди Перлу ,** автору математического аппарата теории сетей Байеса, лауреату премии Тьюринга (2011) за вклад в ИИ посредством разработки исчисления вероятностных рассуждений.

Пример: Сеть тревоги

- Переменные
  - B: Ограбление
  - A: Тревога
  - M: Звонок Мэри
  - J: Звонок Джона
  - E: Землетрясение!





Топология сети показывает, что взлом и землетрясения непосредственно влияют на вероятность появления тревожного сигнала, а звонки Джона и Мэри зависят только от тревожного сигнала. Поэтому сеть подтверждает предположения, что соседи самостоятельно не обнаруживают какие-либо попытки взлома, не замечают незначительных землетрясений и не совещаются друг с другом перед звонками.

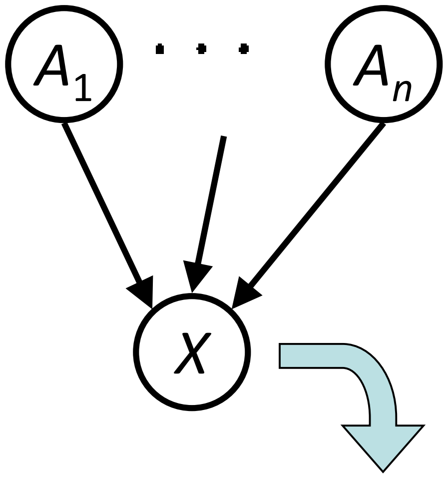
Семантика сетей Байеса

Семантика сетей Байеса

- Множество узлов, каждый узел – одна переменная X
- Направленный, ациклический граф
- Условное распределение для каждого узла:
  - распределение X для каждой комбинации значений родительских вершин

$$P(X|a_1 \ldots a_n)$$

  - распределение представляется в виде таблицы условных вероятностей (CPT- Conditional Probability Table)
  - описывает зашумленные причинные связи



$$P(X|A_1 \ldots A_n)$$

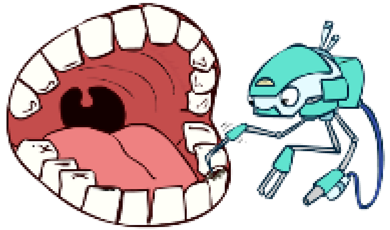
Сеть Байеса= Топология (граф) + Локальные условные вероятности

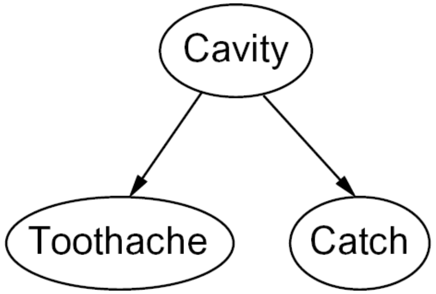
Вероятности в сетях Байеса

- Сети Байеса неявно кодируют совместные распределения:
  - В виде произведения локальных условных распределений;
  - Чтобы определить, какая вероятность соответствует полному присваиванию в BN, перемножьте все соответствующие условные вероятности:

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

  - Пример:





$$P(+cavity, +catch, -toothache) = P(-toothache|+cavity)\ P(+catch|+cavity)\ P(+cavity)$$

Если байесовская сеть служит представлением совместного распределения, то она может также применяться для получения ответа на любой запрос путем суммирования соответствующих элементов совместного распределения.

Вероятности в сетях Байеса

- Почему BN гарантирует правильное вычисление совместного распределения?

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

- Цепочное правило говорит, что :

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1})$$

- Предположив условную независимость:

$$P(x_i | x_1, \ldots x_{i-1}) = P(x_i | parents(X_i))$$

→ получим следствие :

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

Байесовская сеть не только является полным и не избыточным представлением предметной области, но часто оказывается намного более компактной по сравнению с полным совместным распределением. Именно благодаря этому свойству байесовские сети становятся применимыми для представления предметных областей со многими переменными. Компактность байесовских сетей является примером общего свойства **локально структурированных систем.** В локально структурированной системе каждый субкомпонент непосредственно взаимодействует только с ограниченным количеством других компонентов независимо от общего количества компонентов. Локальная структура обычно ассоциируется с линейным, а не с экспоненциальным ростом сложности.

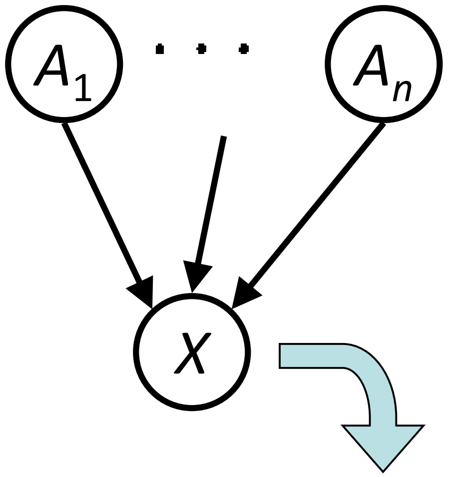
Канонические модели

Существует несколько канонических связей узлов (моделей), используемых в BN:
- Noisy OR и Noisy AND (бинарные переменные);
- Noisy Max и Noisy Min (переменные с несколькими значениями).

Модель Noisy OR основана на концепции логического OR, но отличается наличием определенной (небольшой) вероятности того, что значением следствия не является True, даже, если один или несколько родителей имеют значение True.

Noisy AND связана с логическим оператором AND.

Noisy Max и Noisy Min обеспечивают выбор максимального или минимального значений с учетом вероятности распределения возможных значений следствия.



$$P(X|A_1 \ldots A_n)$$

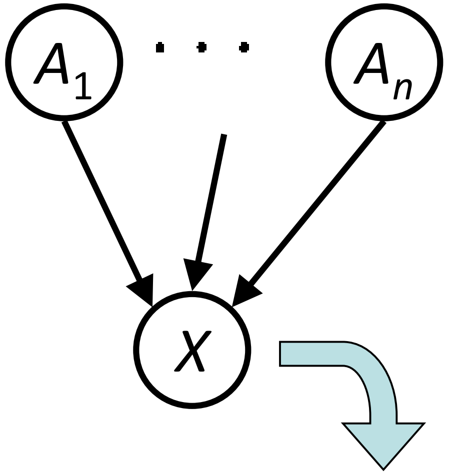
Модель Noisy OR

Модель Noisy OR применяется, когда любая из переменных-предпосылок может иметь значение True и чем больше предпосылок имеют значение True, тем более вырастает вероятность получения именно этого следствия.
Обязательные условия Noisy OR :

- следствие ложное, если все предпосылки *Ai* ложные;
- механизмы подавления (блокирования) следствия по значению одной предпосылки, не зависят от механизмов подавления других предпосылок.

Вероятность подавления следствия *X* при наличии предпосылки *Ai*:

$$Qi=P(X=\text{False} \mid Ai=\text{True})$$



$$P(X|A_1 \ldots A_n)$$

Таблица условных вероятностей для Noisy OR при Q1=Q2=Q3=0.1

A1	0	0	0	0	1	1	1	1
A2	0	0	1	1	0	0	1	1
A3	0	1	0	1	0	1	0	1
P(X=0)	1	0.1	0.1	0.01	0.1	0.01	0.01	0.001
P(X=1)	0	0.9	0.9	0.99	0.9	0.99	0.99	0.999

Пример: Подбрасывание монеты

X_1 X_2 $\cdots$ X_n

$P(X_1)$

h	0.5
t	0.5

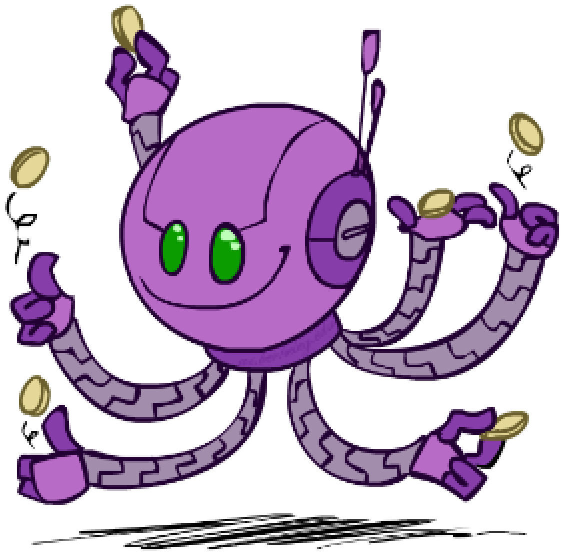
$P(X_2)$

h	0.5
t	0.5

$\cdots$

$P(X_n)$

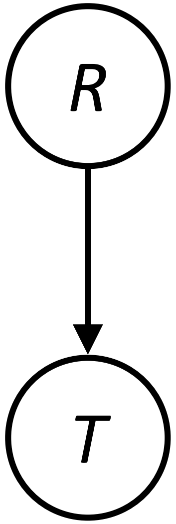
h	0.5
t	0.5



$P(h, h, t, h) =$ P(h)P(h)P(t)P(h)

Только распределения, в которых переменные абсолютно независимы, могут быть представлены сетью Байеса без дуг

Пример: Траффик



$P(R)$

+r	1/4
-r	3/4

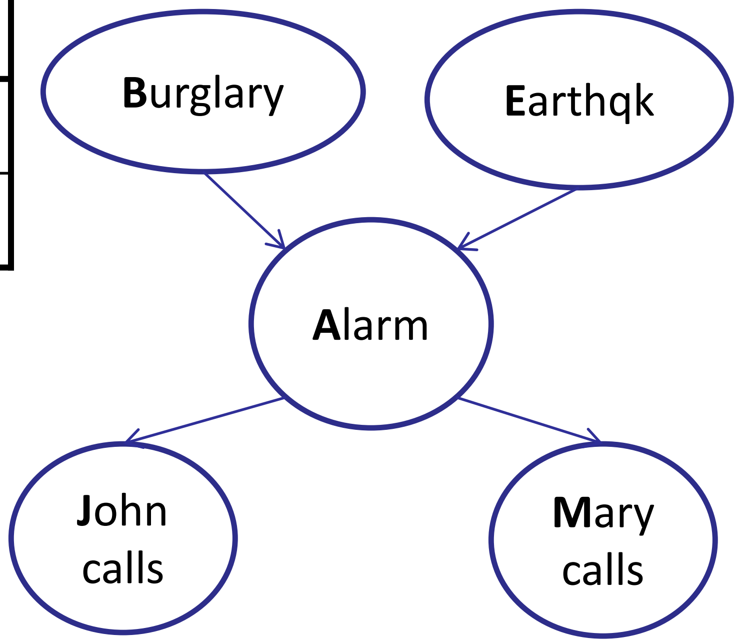
$P(T|R)$

+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

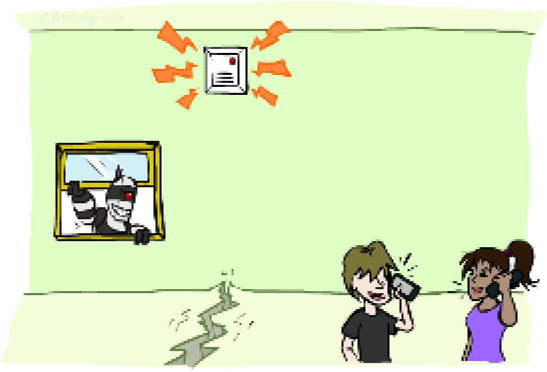
$P(+r, -t) =$ P(+r)P(-t|+r) = ¼*1/4

Пример: Сеть Тревоги

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998



A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

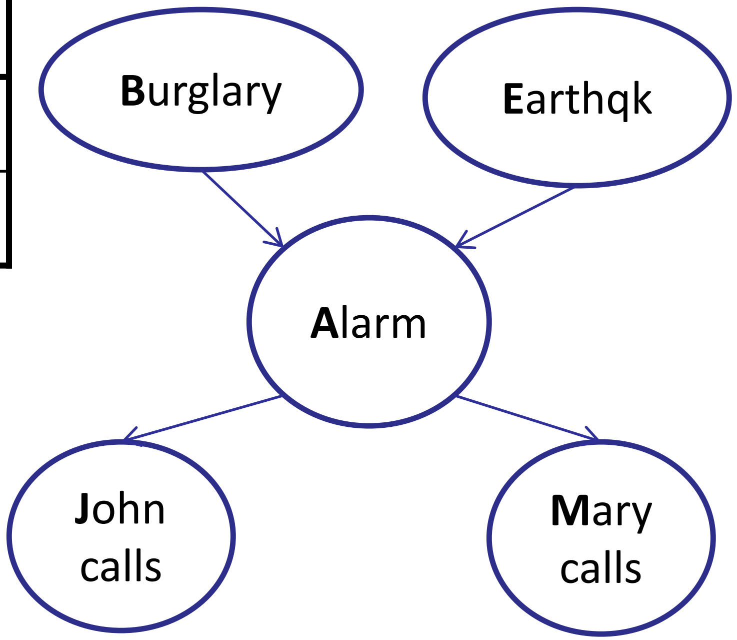
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

P(M|A)P(J|A) P(A|B,E)P(E)P(B) – совместное распределение

Пример: Сеть Тревоги

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

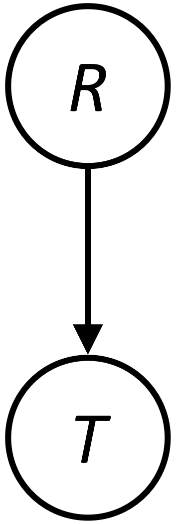
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

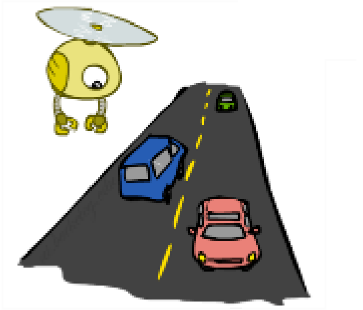
Вероятность того, что прозвучал тревожный сигнал, но не произошли ни взлом, ни землетрясение, а позвонили и Мэри, и Джон.

P(+j ,+m,+a,-Ь,-e)= P(j|a)P(m|a)P(a|-b,-e) P(-b) P(-e)= 0.90 x 0.70 x 0.001 x 0.999 x 0.998 = 0.00062

Пример: Траффик

- Каузальное направление





$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

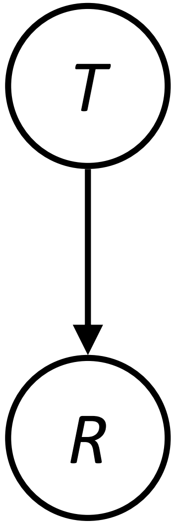
$P(T,R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

P(T,R)=P(T|R) P(R)

Пример: реверс причины

- Реверс каузальности?



$P(T)$

+t	9/16
-t	7/16

$P(R|T)$

+t	+r	1/3
	-r	2/3
-t	+r	1/7
	-r	6/7

P(R|T) =P(T,R)/P(T)

$P(T, R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

Получили тоже самое совместное распределение, что и на предыдущем слайде

Каузальность?

- Если байесовские сети отражают истинные причинные закономерности, то:
  - часто их топология проще (узлы имеют меньше родителей);
  - обычно легче объяснять ход рассуждений;
  - часто бывает проще извлечь знания у экспертов.
- BN, на самом деле, не обязательно должны быть причинными:
  - иногда в домене не существует причинно-следственной сети (особенно, если переменные пропущены);
  - например, рассмотрите переменные Traffic и Drips;
  - в итоге в BN появляются стрелки, отражающие **корреляцию,** а не причинно-следственную связь.
- Что в действительности обозначают стрелки?
  - Топология может отражать причинную структуру;
  - Топология в действительности кодирует условную независимость

$$P(x_i|x_1,\ldots x_{i-1}) = P(x_i|parents(X_i))$$