

Байесовская сеть

Набор данных

Изучите существующие форматы представления Байесовских сетей. Выберите любой из понравившихся вам форматов. Вы можете использовать любую стороннюю библиотеку для чтения (разбора) Байесовских сетей.

Выберите любую Байесовскую сеть. Например их можно найти тут:

- <http://www.bnlearn.com/bnrepository/>
- <http://www.cs.huji.ac.il/~galel/Repository/>

Задача

Реализуйте два алгоритма подсчёта вероятности по Байесовской сети. Они должны искать совместную вероятность комбинации из нескольких событий (не обязательно всех). Например: $P(A = a_i \& B = b_j \& D = d_k)$, где событие C может быть любым.

- Первый алгоритм должен точно вычислять вероятность перебирая все возможные комбинации из значений пропущенных событий.
- Второй алгоритм должен случайно сэмплировать комбинации событий по распределению, которое задаёт Байесовская сеть, и оценивать вероятность заданной комбинации событий частотным методом исходя из числа сэмплированных комбинации, которые содержат заданную.

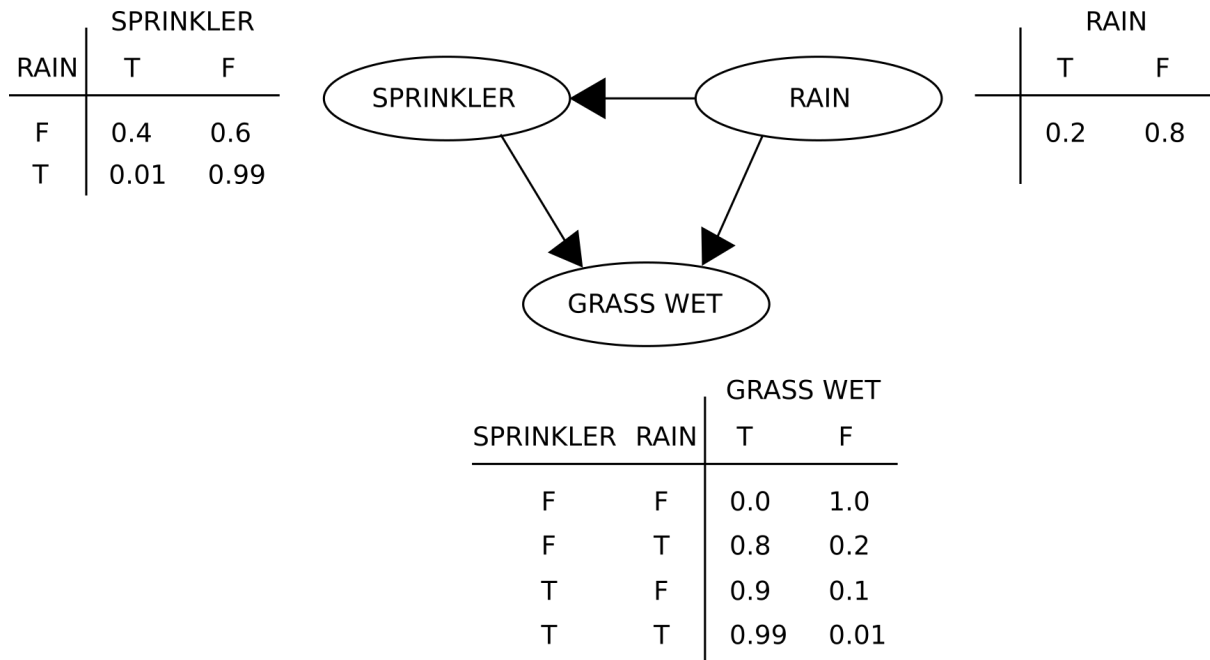
Выберите комбинацию событий для запроса. Сравните значения вероятностей заданной комбинации, полученные первым и вторым алгоритмом. Постройте график зависимости вероятности заданной комбинации, которую оценил второй алгоритм в зависимости от числа сэмплированных комбинаций.

Примечание

Не стоит строго фиксировать набор данных и запрос. Код должен быть написан так, чтобы их можно было в любой момент поменять.

Пример

Рассмотрим простую Байесовскую сеть:



Пусть требуется вычислить вероятность $P(R = F \ \& \ W = T)$.

Тогда первый алгоритм должен найти её как:

$$\begin{aligned} P(R = F \ \& \ W = T) &= P(R = F \ \& \ S = T \ \& \ W = T) + P(R = F \ \& \ S = F \ \& \ W = T) = \\ &= P(R = F) \cdot P(S = T \mid R = F) \cdot P(W = T \mid R = F \ \& \ S = T) + \\ &+ P(R = F) \cdot P(S = F \mid R = F) \cdot P(W = T \mid R = F \ \& \ S = F) = \\ &= 0.8 \cdot 0.4 \cdot 0.9 + 0.8 \cdot 0.6 \cdot 0.0 = 0.288 + 0 = 0.288 \end{aligned}$$

Предположим, что второй алгоритм сэмплировал 3 комбинации:

1. $P(R = F \ \& \ S = F \ \& \ W = F)$
2. $P(R = F \ \& \ S = T \ \& \ W = T)$
3. $P(R = T \ \& \ S = F \ \& \ W = T)$

Тогда искомую комбинацию будет содержать только вторая. Значит согласно второму алгоритму $P(R = F \ \& \ W = T) = \frac{1}{3}$.

Обратите внимание, что различные комбинации могут повторяться. При достаточном числе сэмплированных комбинаций могли бы встретиться редкие комбинации: $P(R = T \ \& \ S = T \ \& \ W = F)$ или $P(R = F \ \& \ S = T \ \& \ W = F)$, однако комбинация $P(R = F \ \& \ S = F \ \& \ W = T)$ не могла встретиться никогда.