Базовые модели принятия решений: байесовские сети

Першин Антон Юрьевич, Ph.D.

Программа «Большие данные и распределенная цифровая платформа»
Санкт-Петербургский государственный университет

Практика по дисциплине «Технологии ИИ»
4 мая 2023 г.

Байесовские сети

- → Байесовские сети являются одной из форм *вероятностных графических моделей*.
- → Они позволяют "закодировать" наше знание о совместном распределении некоторого процесса в форме графа.
- → Под "знанием" здесь в первую очередь подразумевается понимание того, какие случайные величины зависят друг от друга, а какие являются независимыми.
- → Байесовская сеть описывается ориентированным ациклическим графом (directed acyclic graph, DAG):
  - Каждый узел сети соответствует переменной.
  - Направленные ребра указывают на прямые вероятностные отношения; если мы придаем направлению ребер смысл причинно-следственных связей, то такая байесовская сеть называется причинно-следственной сетью (causal network).
  - С каждым узлом X_i связано условное распделение $P(X_i|\mathrm{Pa}(X_i))$, где $\mathrm{Pa}(X_i)$ представляет родителей X_i в графе.

Пример байесовской сети

- → Рассмотрим байесовскую сеть для задачи мониторинга состояния спутников.
- → Каждая переменная здесь является бинарной.
- → Отказ батареи или выход из строя солнечной панели может привести к отказу системы электропитания.
- → Отказ системы электропитания может привести к отклонению траектории, а также к потери связи.

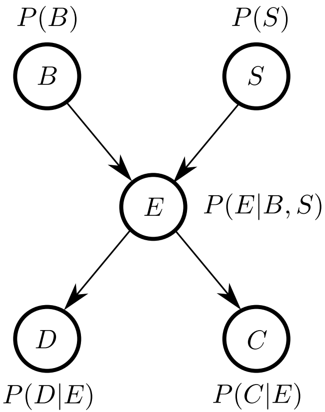

Рис.: B – отказ батареи, S – отказ солнечной панели, E – отказ системы электропитания, D – отклонение от траектории, C – потеря связи.

Цепное правило для байесовской сети

→ Цепное правило для байесовской сети определяет построение совместного распределения из локальных распределений условной вероятности.

→ Пусть в байесовской сети имеются переменные $X_1, \ldots, X_n$. Тогда вероятность присвоения им конкретных значений $x_1, \ldots, x_n$ имеет вид

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | pa(x_i)), \tag{1}$$

где $pa(x_i)$ обозначает присвоение конкретного значения родителю X_i.

→ В нашем примере мы можем захотеть определить вероятность того, что все в порядке, то есть $b = s = e = d = c = 0$:

$$\begin{aligned} P(b = 0, s = 0, e = 0, d = 0, c = 0) = \\ & P(b = 0)P(s = 0)\times \\ & P(e = 0 | b = 0, s = 0)\times \\ & P(d = 0 | e = 0)P(c = 0 | e = 0) \end{aligned} \tag{2}$$

→ Преимущество байесовской сети состоит в том, что без нее нам бы потребовался $2^5 - 1 = 31$ параметр для задания совместного распределения, а с ней – только $1 + 1 + 4 + 2 + 2 = 10$

Условная независимость и d-разделение

- Под условной независимостью $X \perp Y|Z$ понимается $P(X,Y|Z) = P(X|Z)P(Y|Z)$, из чего следует $X \perp Y|Z \Longrightarrow P(X|Z,Y) = P(X|Z)$.
- В нашем пример $D \perp C|E$: если известно, что произошел сбой питания, информация о потери связи не влияет на вероятность наличия отклонения от траектории.
- Байесовская сеть позволяет определить, являются ли две переменные A, B условно независимыми при заданных обусловливающих переменных, используя d-разделение.

Наивный байесовский классификатор

- *Наивный байесовский классификатор* является простейшим примером байесовской сети.
- Он называется “наивным”, потому что он предполагает условную независимость признаков при данном значении класса.

Пусть y обозначает класс, а $x_1, \ldots, x_n$ признаки. По формуле Байеса:

$$P(y|x_1, \ldots, x_n) = \frac{P(x_1, \ldots, x_n|y)P(y)}{P(x_1, \ldots, x_n)}. \tag{3}$$

Условная независимость признаков подразумевает

$$P(x_i|y, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = P(x_i|y). \tag{4}$$

Тогда имеем

$$P(x_1, \ldots, x_n|y) = P(x_2, \ldots, x_n|y, x_1)P(x_1|y) = P(x_2, \ldots, x_n|y)P(x_1|y) = \cdots = \prod_{i=1}^{n} P(x_i|y), \tag{5}$$

что дает

$$P(y|x_1, \ldots, x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1, \ldots, x_n)}. \tag{6}$$

Наивный байесовский классификатор

Наивный байесовский классификатор далее рассматривает оценку апостериорного максимума:

$$\hat{y} = \operatorname{argmax}_y P(y) \prod_{i=1}^{n} P(x_i|y). \tag{7}$$

Для построения классификатора необходимо выбрать априорное распределение $P(y)$ и выбрать параметрическую форму для распределений признаков $P(x_i|y)$.

→ Пусть $\mathcal{D} = \{(\boldsymbol{x}^{(i)}, y_i)\}_{i=1}^{n}$

→ Априорное распределение: $P(y) = \dfrac{|\{(\boldsymbol{x}^{(i)}, y_i) \in \mathcal{D} \mid y_i = y\}|}{|\mathcal{D}|}$

→ Распределение признаков:

- Гауссовский наивный Байес: $P(x_i|y) \sim \mathrm{n}(\mu_y, \sigma_y^2)$, где μ_y и σ_y^2 оценивается по максимальному правдоподобию.
- Мультиномиальный наивный Байес: предлагается, что все признаки являются частотными, то есть x_i обозначает, сколько раз произошло событие i в данном элементе выборки. Тогда $p(x_1, \ldots, x_n) = \frac{\sum_{i=1}^{n} x_i!}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} p_{ki}^{x_i}$, где по максимальному правдоподобию можно оценить p_{ki}. Используется для классификации документов (метод "мешка слов", bag-of-words)