ГРАФИЧЕСКИЕ ВЕРОЯТНОСТНЫЕ МОДЕЛИ

Сергей Николенко

СПбГУ — Санкт-Петербург

25 ноября 2020 г.

Random facts:

- 25 ноября 1120 г. у берегов Нормандии потерпел крушение «Белый корабль»; погибли многие англонормандские аристократы, в том числе наследник престола Вильгельм Аделин; Генрих I провозгласил наследницей свою дочь Матильду, аристократии это не понравилось, и после смерти Генриха в 1135 году началась гражданская война
- 25 ноября 1343 г. цунами, вызванное землетрясением в Тирренском море, разрушило республику Амальфи; вскоре она была поглощена герцогством Салерно
- 25 ноября 1867 г. Альфред Нобель запатентовал динамит, а 25 ноября 1946 г. был основан МФТИ
- 25 ноября 1970 г. Юкио Мисима, находясь с официальным визитом на базе войск сил самообороны в Итигае, взял в заложники командующего базой и обратился к солдатам с призывом совершить государственный переворот; но солдаты его по большей части проигнорировали, и Мисима совершил харакири

Графические вероятностные модели

В чём же проблема

- В предыдущих лекциях мы рассмотрели задачу байесовского вывода, ввели понятие сопряжённого априорного распределения, поняли, что наша основная задача – найти апостериорное распределение.
- Но если всё так просто – взяли интеграл, посчитали, всё получилось – о чём же здесь целая наука?
- Проблема заключается в том, что распределения, которые нас интересуют, обычно слишком сложные (слишком много переменных, сложные связи).
- Но, с другой стороны, в них есть дополнительная структура, которую можно использовать, структура в виде независимостей и условных независимостей некоторых переменных.

Графические модели с направленными графами

- Пример: рассмотрим распределение трёх переменных и запишем его по формуле полной вероятности:

$$p(x, y, z) = p(x \mid y, z) p(y \mid z) p(z).$$

- Теперь нарисуем граф, в котором стрелки указывают, какие условные вероятности заданы.
- Пока граф полносвязный, это нам ничего не даёт – любое распределение $p(x_1, \dots, x_n)$ так можно переписать.
- Но если некоторых связей *нет*, это даёт нам важную информацию и упрощает жизнь.

- Рассмотрим направленный ациклический граф на вершинах $x_1, \ldots, x_k$ и зададим в каждой вершине распределения $p(x_i \mid \mathrm{pa}(x_i))$. Тогда будем говорить, что граф с этими локальными распределениями является графической моделью (байесовской сетью доверия) для совместного распределения вероятностей

$$p(x_1, \ldots, x_k) = \prod_{i=1}^{k} p(x_i \mid \mathrm{pa}(x_i)).$$

- Другими словами, если мы можем разложить большое совместное распределение в произведение локальных распределений, каждое из которых связывает мало переменных, это хорошо. :)

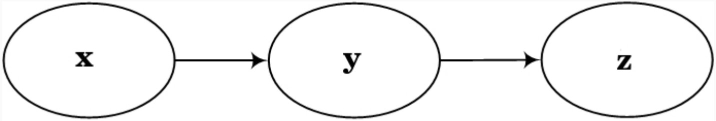
Графические модели с направленными графами

- Пример: обучение параметров распределения по нескольким экспериментам (плашки, можно нарисовать параметры явно):

$$p(x_1, \dots, x_n, \theta) = p(\theta) \prod_{i=1}^{n} p(x_i \mid \theta).$$

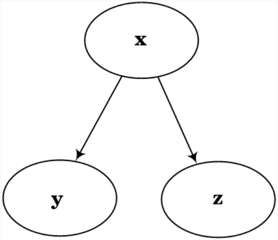
- Что можно сказать о (не)зависимости случайных величин x_i и x_j?
- Задача вывода на графической модели: в некоторой части вершин значения наблюдаются, надо пересчитать распределения в других вершинах (подсчитать условные распределения). Например, из этой модели получатся и задача обучения параметров, и задача последующего предсказания.

- d-разделимость – условная независимость, выраженная в структуре графа:
  - последовательная связь, $p(x, y, z) = p(x)p(y \mid x)p(z \mid y)$:
    - если y не наблюдается, то $p(x, z) = p(x) \int p(y \mid x)p(z \mid y)\mathrm{d}y = p(x)p(z \mid x)$;
    - если y наблюдается, то $p(x, z \mid y) = \frac{p(x,y,z)}{p(y)} = \frac{p(x)p(y|x)p(z|y)}{p(y)} = p(x \mid y)p(z \mid y)$, получили условную независимость.



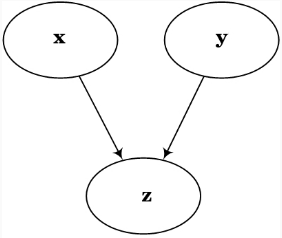
Последовательная связь

- расходящаяся связь, $p(x, y, z) = p(x)p(y \mid x)p(z \mid x)$, – так же:
    - если y не наблюдается, то $p(x, z) = p(x)p(z \mid x) \int p(y \mid x)\mathrm{d}y = p(x)p(z \mid x)$;
    - если y наблюдается, то $p(x, z \mid y) = \frac{p(x,y,z)}{p(y)} = \frac{p(x)p(y|x)p(z|x)}{p(y)} = p(x \mid y)p(z \mid y)$, получили условную независимость.



Расходящаяся связь

- Интересный случай – сходящаяся связь, $p(x, y, z) = p(x)p(y)p(z \mid x, y)$:
  - если z не наблюдается, то $p(x, y) = p(x)p(y)$, независимость есть;
  - если z наблюдается, то $p(x, y \mid z) = \frac{p(x,y,z)}{p(z)} = \frac{p(x)p(y)p(z|x,y)}{p(z)}$, и условной независимости нету.



Сходящаяся связь

Обобщение: если наблюдается хотя бы один из потомков z, уже может не быть независимости между x и y.

Графические модели с направленными графами

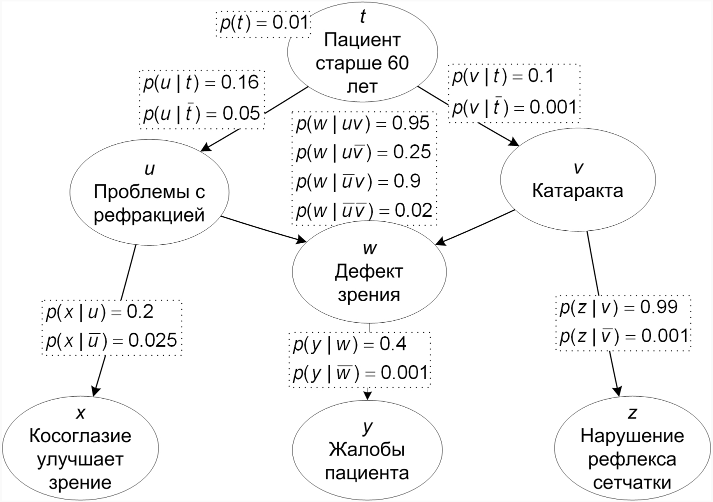
- Можно сформулировать, как структура графа соотносится с условной независимостью: в графе, где вершины из множества Z получили означивания (evidence), две ещё не означенные вершины x и y условно независимы при условии множества означенных вершин Z, если любой (ненаправленный) путь между x и y:
  - либо проходит через означенную вершину $z \in Z$ с последовательной или расходящейся связью;
  - либо проходит через вершину со сходящейся связью, в которой ни она, ни её потомки не получили означиваний.

Графические модели с направленными графами

- Можно сказать, что граф задаёт некоторое семейство распределений – не все распределения на вершинах графа будут соответствовать тем ограничениям по условной независимости, которые накладывает структура графа.
- Теорема (без доказательства): это семейство распределений в точности совпадает с семейством тех распределений, которые можно разложить в произведение

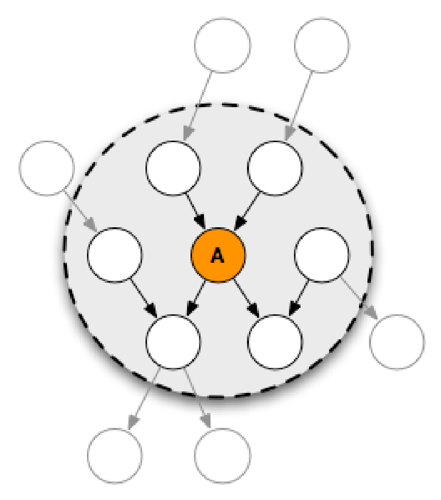
$$p(x_1, \ldots, x_k) = \prod_{i=1}^{k} p(x_i \mid \mathrm{pa}(x_i)).$$

Пример байесовской сети



- Интересный вопрос: какие вершины нужно означить, чтобы наверняка «отрезать» одну вершину (Markov blanket)?
- Иначе говоря, для какого минимального множества вершин X $p(x_i \mid x_{j\neq i}) = p(x_i \mid X)$?

Markov blanket

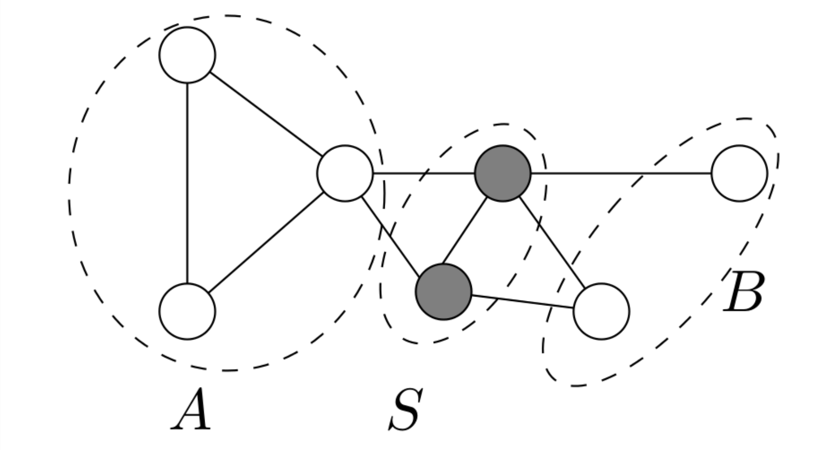


Другие графические модели

Графические модели с ненаправленными графами

- Можно сделать и так, чтобы условие независимости было (более) локальным.
- Для этого нужно задавать модели ненаправленными графами. В них условие совсем естественное: множество вершин X условно независимо от множества вершин Y при условии множества вершин Z, если любой путь от X к Y проходит через Z.
- В частности, очевидно, $p(x_i, x_j \mid x_{k\neq i,j}) = p(x_i \mid x_{k\neq i,j})p(x_j \mid x_{k\neq i,j})$ тогда и только тогда, когда x_i и x_j не соединены ребром.
- Такие модели называются *марковскими сетями* (Markov random fields).

Условная независимость в ненаправленных моделях



- Поэтому в ненаправленных моделях локальные распределения соответствуют кликам в графе, и факторизация получается в виде

$$p(x_1, \ldots, x_k) = \frac{1}{Z} \prod \psi_C(x_C),$$

  где C – максимальные клики, ψ_C – неотрицательные функции (*потенциалы*), а Z – нормировочная константа (partition function).
- Поскольку $\psi_C \geq 0$, их обычно представляют как экспоненты:

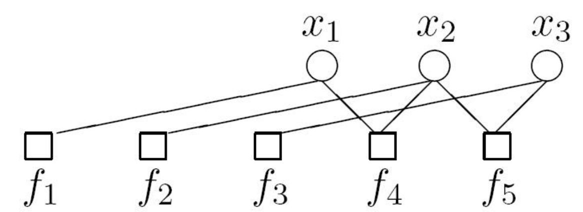
$$\psi_C(x_C) = \exp\left(-E_C(x_C)\right),$$

  E_C – функции энергии, они суммируются в полную энергию системы (это всё похоже на статистическую физику, отсюда и терминология).

- Интересный факт: назовём *идеальной картой* (perfect map) распределения D графическую модель G, если все условные независимости, присутствующие в D, отображены в G, и наоборот (ничего лишнего). Тогда идеальные карты в виде направленных моделей существуют не у всех распределений, в виде ненаправленных тоже не у всех, и эти множества существенно различаются (бывают распределения, которые нельзя идеально выразить направленной моделью, но можно ненаправленной, и наоборот).

Фактор-графы

- Важная для вывода модификация – *фактор-граф* (можно построить и по направленной модели, и по ненаправленной).
- Фактор-граф – двудольный граф функций и переменных.
- Функция, соответствующая графу, – произведение всех входящих в него функций (т.е. то самое разложение и есть).
- Пример: $p(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)f_3(x_3)f_4(x_1, x_2)f_5(x_2, x_3)$.



Спасибо!

Спасибо за внимание!