

Байесовское мультимоделирование: графические модели

Московский Физико-Технический Институт

2021

Графические модели

Условная независимость

События X, Y называются условно независимыми при условии Z : $X \perp Y|Z$, если

$$P(X|Y, Z) = P(X|Z).$$

Условная зависимость

События X, Y называются условно зависимыми при условии множества всех событий $\mathfrak{G} : X, Y \in \mathfrak{G}$, если

$$X \not\perp Y|\mathfrak{G} \setminus \{X, Y\}.$$

Графическая модель

Вероятностная модель называется графической, если ее можно представить в виде графа, где ребро ставится между условно зависимыми событиями.

Не графические модели

Не являются графическими моделями:

- Многослойные перцептроны (в общем виде), деревья решений и пр.
- Ненаправленные модели со сложным взаимодействием (есть взаимодействие на уровне подмножества вершин, но нет взаимодействия на уровне клик).

Виды графических моделей

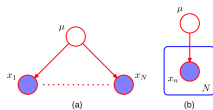
- Направленные (байесовские сети)
 - ▶ Удобны для проектирования моделей
- Ненаправленные (марковские сети)
- Фактор-графы
 - ▶ Удобны для вывода и оптимизации

Байесовские сети

- Задается направленным графом без циклов
- Совместное распределение для графа K вершин:

$$p(v_1, \dots, v_K) = \prod_{i=1}^K p(v_i | \text{parent}(v_i))$$

- Пример: линейная регрессия



DAG и Plate notation (Bishop)

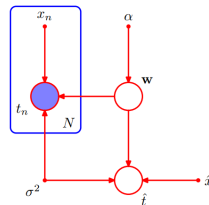


Plate notation для модели регрессии (Bishop)

Элементы причинного графа

$X \rightarrow Y \rightarrow Z$ — цепочка

Пример:

- X — бюджет школы
- Y — средний балл учеников
- Z — доля поступающих в ВУЗы

Свойства:

- 1 X и Y , Y и Z — зависимы:
 $\exists x, y : P(Y = y | X = x) \neq p(Y = y)$
 $\exists y, z : P(Z = z | Y = y) \neq p(Z = z)$
- 2 Z и X — скорее всего, зависимы
- 3 $Z \perp X | Y$ — условно независимы: $\forall x, y, z$

$$P(Z = z | X = x, Y = y) = P(Z = z | Y = y)$$

(если Y фиксировано, то X и Z независимы)

Элементы причинного графа

$$X \leftarrow Y \rightarrow Z \text{ — вилка}$$

Пример:

- X — продажи мороженого
- Y — средняя дневная температура воздуха
- Z — число преступлений

Свойства:

- ① X и Y , Y и Z — зависимы
- ② X и Z — скорее всего, зависимы
- ③ $X \perp Z | Y$ — условно независимы

Элементы причинного графа

$$Y \rightarrow X \leftarrow Z \text{ — коллайдер}$$

Пример (заболевание вирусом):

- X — осложнения
- Y — возраст
- Z — хронические болезни

Свойства:

- 1 Y и X , Z и X — зависимы
- 2 Y и Z — независимы
- 3 $Y \not\perp Z | X$ — условно зависимы

d-разделимость

Путь P блокируется переменной Z , если:

- 1 P содержит $A \rightarrow B \rightarrow C$, $A \leftarrow B \rightarrow C$, $B \in Z$
- 2 P содержит $A \rightarrow B \leftarrow C$, $B \notin Z$ и все потомки $B \notin Z$

Если Z блокирует все пути из X в Y , то X и Y **d-разделимы**:

$$X \perp Y | Z.$$

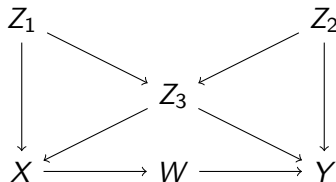
d-разделимость

Путь P блокируется переменной Z , если:

- 1 P содержит $A \rightarrow B \rightarrow C$, $A \leftarrow B \rightarrow C$, $B \in Z$
- 2 P содержит $A \rightarrow B \leftarrow C$, $B \notin Z$ и все потомки $B \notin Z$

Если Z блокирует все пути из X в Y , то X и Y d-разделимы.

Пример:



Упорядоченная пара вершин	d-разделяющее множество
(Z_1, W)	X

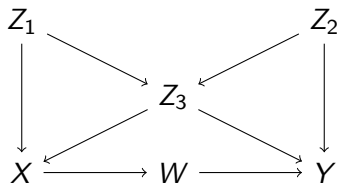
d-разделимость

Путь P блокируется переменной Z , если:

- 1 P содержит $A \rightarrow B \rightarrow C$, $A \leftarrow B \rightarrow C$, $B \in Z$
- 2 P содержит $A \rightarrow B \leftarrow C$, $B \notin Z$ и все потомки $B \notin Z$

Если Z блокирует все пути из X в Y , то X и Y d-разделимы.

Пример:



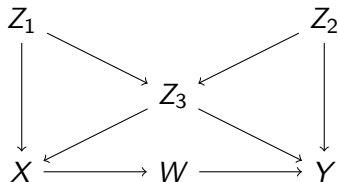
Упорядоченная пара вершин	d-разделяющее множество
(Z_1, W)	X
(Z_1, Y)	$\{Z_3, X, Z_2\}, \{Z_3, W, Z_2\}$

d-разделимость

Путь P блокируется переменной Z , если:

- 1 P содержит $A \rightarrow B \rightarrow C$, $A \leftarrow B \rightarrow C$, $B \in Z$
- 2 P содержит $A \rightarrow B \leftarrow C$, $B \notin Z$ и все потомки $B \notin Z$

Если Z блокирует все пути из X в Y , то X и Y d-разделимы.



Упорядоченная пара вершин	d-разделяющее множество
(Z_1, W)	X
(Z_1, Y)	$\{Z_3, X, Z_2\}, \{Z_3, W, Z_2\}$
(X, Y)	$\{W, Z_3, Z_1\}$

Марковские случайные поля

Модели представимы в виде ненаправленного графа.

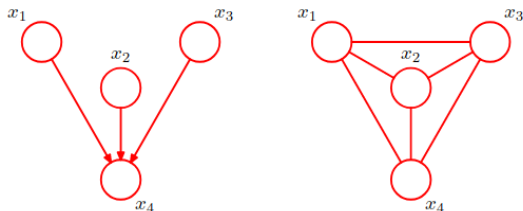
Отличия от байесовских сетей:

- Нет направленности \rightarrow нельзя выявить причинность события, только наличие взаимодействия.
- Правдоподобие факторизуется следующим образом:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_C \psi(\mathbf{x}_C),$$

где \mathbf{x}_C — максимальная клика на графе, $\psi \geq 0$ — потенциальная функция.

- Условная независимость: если все пути от A до B проходят через C , то $A \perp B | C$.



Вывод в цепочках



Наивный подсчет правдоподобия для x_n :

$$p(x_n) = \sum_{x_1} \sum_{x_2} \dots, \sum_{x_{n-1}} \sum_{x_{n+1}} \dots \sum_{x_N} p(\mathbf{x}),$$

Для N дискретных переменных с K состояниями сложность: $O(K^N)$

Вывод в цепочках: перегруппировка слагаемых

$$p(x_n) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} p(\mathbf{x}),$$

$$p(\mathbf{x}) = \psi(x_1, x_2) \psi(x_2, x_3) \cdots \psi(x_{N-1}, x_N).$$

Перегруппируем слагаемые:

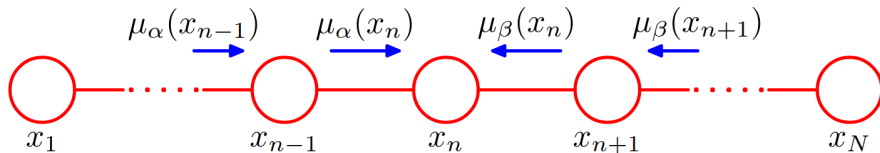
$$\begin{aligned} p(x_n) = & \sum_{x_{n-1}} \psi(x_{n-1}, x_n) \cdots \left(\sum_{x_1} \psi(x_1, x_2) \right) \times \\ & \times \left(\sum_{x_{n+1}} \psi(x_n, x_{n+1}) \cdots \left(\sum_{x_N} \psi(x_{N-1}, x_N) \right) \right). \end{aligned}$$

За счет группировку слагаемых и “скобочек” сложность: $O(NK^2)$.

Message passing

$$p(x_n) = \underbrace{\sum_{x_{n-1}} \psi(x_{n-1}, x_n) \dots \left(\sum_{x_1} \psi(x_1, x_2) \right)}_{\mu_a(x_n)} \times \underbrace{\left(\sum_{x_{n+1}} \psi(x_n, x_{n+1}) \dots \left(\sum_{x_N} \psi(x_{N-1}, x_N) \right) \right)}_{\mu_b(x_n)}.$$

Интерпретация: $\mu_a(x_n)$ — сообщение, переданное вперед от x_{n-1} до x_n , $\mu_b(x_n)$ — сообщение, переданное назад от x_{n+1} .



Вывод в цепочках: детали

Вывод представим в виде итеративного алгоритма:

- вычисляем вектор $\sum_{x_1} \psi(x_1, x_2) = \mu_a(x_2)$, хранящий значение $\mu_a(x_2)$ для всех значений переменной x_2 ;
- вычисляем вектор $\sum_{x_2} \psi(x_2, x_3) (\sum_{x_1} \psi(x_1, x_2)) = \sum_{x_2} \psi(x_2, x_3) \mu_a(x_2) = \mu_a(x_3)$;
- ...
- вычисляем вектор $\sum_{x_{n+1}} \psi(x_n, x_{n+1}) \mu_b(x_{n+1}) = \mu_b(x_n)$.
- для направленных моделей, где

$$\psi(x_1, x_2) = p(x_1)p(x_2|x_1), \quad \psi(x_i, x_{i+1}) = p(x_{i+1}|x_i),$$

вычислять μ_b не требуется:

$$\begin{aligned} \mu_b(x_n) &= \sum_{x_{n+1}} \psi(x_n, x_{n+1}) \dots \left(\sum_{x_N} \psi(x_{N-1}, x_N) \right) = \\ &= \sum_{x_{n+1}} p(x_{n+1}|x_n) \dots \left(\sum_{x_N} p(x_N|x_{N-1}) \right) = 1. \end{aligned}$$

Фактор-графы

Определение

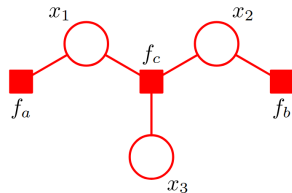
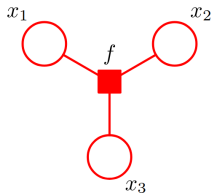
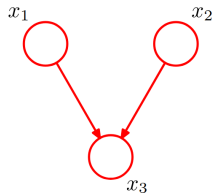
Фактор-граф — это двудольный граф с двумя типами вершин: переменными и факторами.

Правдоподобие определяется как произведение факторов:

$$p(\mathbf{x}) = \prod_i f_i.$$

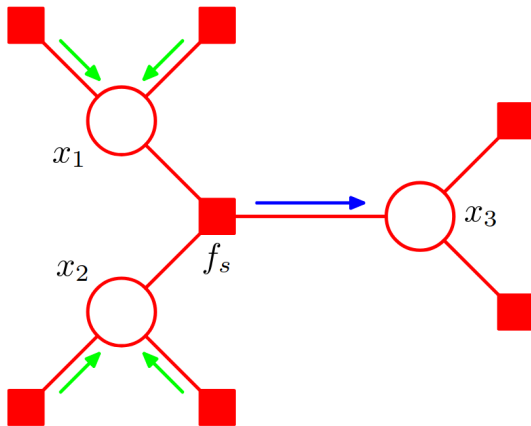
Пример: модель $p(x_1)p(x_2)p(x_3|x_2, x_1)$ и два варианта факторизации:

$$f = p(x_1)p(x_2)p(x_3|x_2, x_1), \quad f_a = p(x_1), f_b = p(x_2), f_c = p(x_1)p(x_2)p(x_3|x_2, x_1).$$



Вывод в фактор-графах: иллюстрация

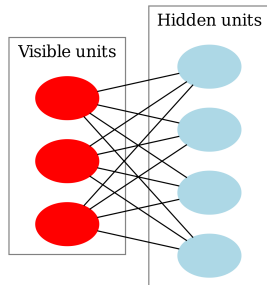
Алгоритм sum-product: правдоподобие расписывается через композицию сообщений, получаемых от факторов до переменных вершин.



Пример моделей: RBM

$$p(\mathbf{x}, \mathbf{h}) = \frac{1}{Z} \exp(-E(\mathbf{x}, \mathbf{h})),$$

$$E = -\mathbf{w}_1^\top \mathbf{x} - \mathbf{w}_2^\top \mathbf{h} - \mathbf{x}^\top \mathbf{W}_3 \mathbf{h}.$$



Пример моделей: Structured VAEs

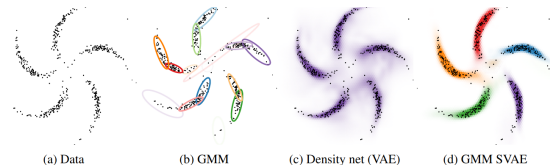
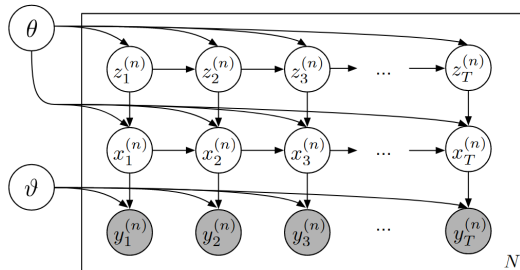
В основе модели SLDS:

$$z_{t+1}|z_t \sim \pi^{t+1},$$

$$y_t \sim \mathcal{N}(\text{MLP}^{z_t}(x_t)).$$

Оптимизация — оптимизация вариационной оценки правдоподобия.

При выводе и оптимизации используется message passing.



Литература и прочие ресурсы

- Bishop C. M. Pattern recognition //Machine learning. – 2006. – Т. 128. – №. 9.
- Edwards D. Introduction to graphical modelling. – Springer Science & Business Media, 2012.
- Pearl J., Glymour M., Jewell N. P. Causal inference in statistics: A primer. – John Wiley & Sons, 2016.
- Hinton G. E., Salakhutdinov R. R. Reducing the dimensionality of data with neural networks //science. – 2006. – Т. 313. – №. 5786. – С. 504-507.
- https://en.wikipedia.org/wiki/Restricted_Boltzmann_machine
- Johnson M. J. et al. Structured VAEs: Composing probabilistic graphical models and variational autoencoders //arXiv preprint arXiv:1603.06277. – 2016. – Т. 2. – С. 2016.
- Johnson M. J. et al. Composing graphical models with neural networks for structured representations and fast inference //Advances in neural information processing systems. – 2016. – Т. 29. – С. 2946-2954.

Перекрестные доклады по лабам

- Время на доклад: до двух минут
- Нужно донести:
 - ▶ **Какая задача решалась в лабе**
 - ▶ Как это было решено (если есть какие-то технические особенности)
 - ▶ **Какие графики получены и как их интерпретировать**
- Лабораторные для докладов будут распределены случайно (см. гитхаб)