Байесовское мультимоделирование: графические модели

Московский Физико-Технический Институт

2021

Графические модели

Условная независимость

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

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

Условная зависимость

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

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

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 взаимодейсвтие на уровне подмножества вершин, но нет взаимодейсвтия на уровне клик).

Виды графических моделей

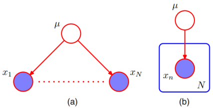
- Направленные (байесовские сети)
  - Удобны для проектирования моделей
- Ненаправленные (марковские сети)
- Фактор-графы
  - Удобны для вывода и оптимизации

Байесовские сети

- Задается направленным графом без циклов
- Совметное 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)

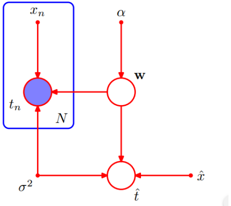


Plate notation для модели регрессии (Bishop)

Элементы причинного графа

$$X \to Y \to Z \text{ — } \textbf{цепочка}$$

Пример:

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

Свойства:

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

$$\boldsymbol{P}(Z = z \,|X = x, Y = y\,) = \boldsymbol{P}(Z = z \,|Y = y\,)$$

(если Y фиксировано, то X и Z независимы)

Элементы причинного графа

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

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

$$Y \to X \leftarrow Z \text{ — } \textbf{коллайдер}$$

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 $A \to B \to C$, $A \leftarrow B \to C$, $B \in Z$
2. P содержит $A \to B \leftarrow C$, $B \notin Z$ 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 $A \to B \to C$, $A \leftarrow B \to C$, $B \in Z$
2. P содержит $A \to B \leftarrow C$, $B \notin Z$ и все потомки $B \notin Z$

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

Пример:

$$\begin{tikzcd}
Z_1 \arrow[dr] \arrow[dd] & & Z_2 \arrow[dl] \arrow[dd] \\
& Z_3 \arrow[dl] \arrow[dr] & \\
X \arrow[r] & W \arrow[r] & Y
\end{tikzcd}$$

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

d-разделимость

Путь P **блокируется** переменной Z, если:

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

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

Пример:

$$\begin{tikzcd}
Z_1 \arrow[dd] \arrow[dr] & & Z_2 \arrow[dl] \arrow[dd] \\
& Z_3 \arrow[dl] \arrow[dr] & \\
X \arrow[r] & W \arrow[r] & Y
\end{tikzcd}$$

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 $A \to B \to C$, $A \leftarrow B \to C$, $B \in Z$
2. P содержит $A \to B \leftarrow C$, $B \notin Z$ и все потомки $B \notin Z$

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

$$\begin{tikzcd}
Z_1 \arrow[dd] \arrow[dr] & & Z_2 \arrow[dl] \arrow[dd] \\
& Z_3 \arrow[dl] \arrow[dr] & \\
X \arrow[r] & W \arrow[r] & Y
\end{tikzcd}$$

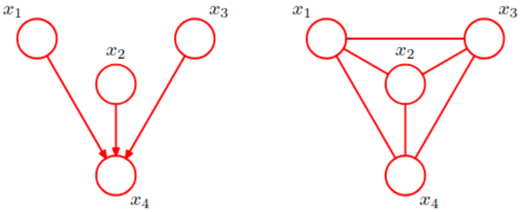
Упорядоченная пара вершин	d-разделяющее множество
(Z_1, W)	X
(Z_1, Y)	$\{Z_3, X, Z_2\}, \{Z_3, W, Z_2\}$
(X, Y)	$\{W, Z_3, Z_1\}$

Марковские случайные поля

Модели представимы в виде ненаправленного графа.

Отличия от байесовских сетей:

- Нет направленности $\to$ нельзя выявить причинность события, только наличие взаимодействия.
- Правдоподобие факторизуется следующим образом:
$$p(\boldsymbol{x}) = \frac{1}{Z}\prod_{C}\psi(\boldsymbol{X}_C),$$
где $\boldsymbol{X}_C$ — 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

(Bishop)

Наивный подсчет правдоподобия для x_n:

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

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

Вывод в цепочках: перегруппировка слагаемых

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

$$p(\boldsymbol{x}) = p(x_1)p(x_2|x_1)p(x_3|x_2)\ldots p(x_N|X_{N-1}).$$

Перегруппируем слагаемые:

$$p(x_n) = \sum_{x_{n-1}} p(x_n|x_{n-1})\ldots\Big(\sum_{x_2} p(x_2|X_1)\Big(\sum_{x_1} p(x_1)\Big)\Big)\times$$

$$\times\Big(\sum_{x_{n+1}} p(x_{n+1}|X_n)\ldots\Big(\sum_{x_N} p(x_N|x_{N-1})\Big)\Big).$$

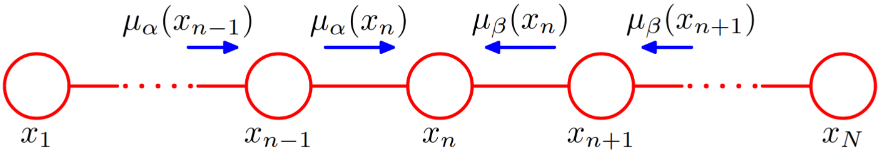
За счет группировку слагаемых и “скобочек” сложность: $O(NK^2)$.

Message passing

$$p(x_n) = \underbrace{\sum_{x_{n-1}} p(x_n|x_{n-1}) \dots \Big(\sum_{x_2} p(x_2|X_1)\Big(\sum_{x_1} p(x_1)\Big)\Big)}_{\mu_a(x_n)} \times$$

$$\times \underbrace{\Big(\sum_{x_{n+1}} p(x_{n+1}|X_n) \dots \Big(\sum_{x_N} p(x_N|x_{N-1})\Big)\Big)}_{\mu_b(x_n)} .$$

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}.



Фактор-графы

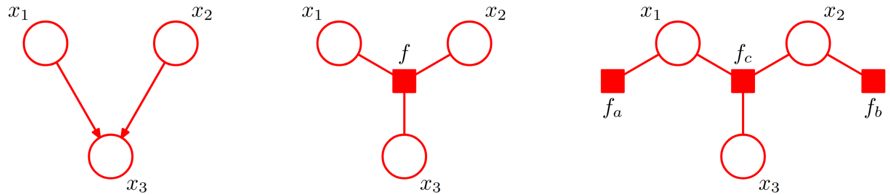
Определение

Фактор-граф — это двудольный граф с двумя типами вершин: переменными и факторами.
Правдоподобие определяется как произведение факторов:

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

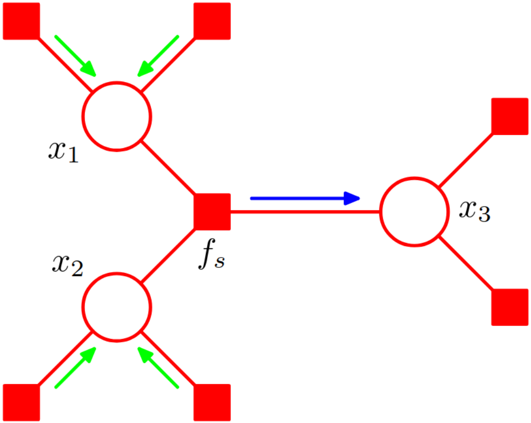
Пример: модель $p(x_1)p(x_2)p(x_3|x_2,x_1)$ и два варианта факторизации:

$$f = p(x_1)p(x_2)p(x_3|x_2,x_1), \quad f_a = p(x_1), f_b = p(x_2), f_3 = p(x_1)p(x_2)p(x_3|x_2,x_1.$$



Вывод в фактор-графах: иллюстрация

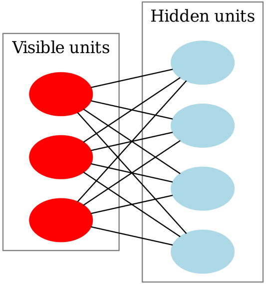
Алгоритм sum-product: правдоподобие расписывается через композицию сообщений, получаемых от факторов до переменных вершин.



Пример моделей: RBM

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

$$E = -\boldsymbol{w}_1^{\mathsf{T}}\boldsymbol{x} - \boldsymbol{w}_2^{\mathsf{T}}\boldsymbol{h} - \boldsymbol{x}^{\mathsf{T}}\boldsymbol{W}_3\boldsymbol{h}.$$



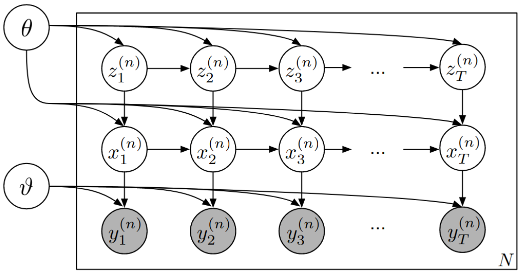
Пример моделей: Structured VAEs

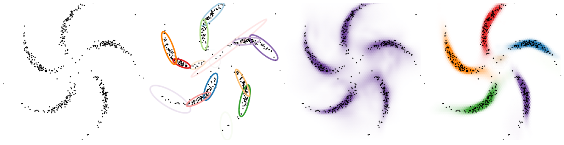
В основе модели SLDS:

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

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

Оптимизация — оптимизация вариационной оценки правдоподобия.
При выводе и оптимизации используется message passing.





(a) Data (b) GMM (c) Density net (VAE) (d) GMM SVAE

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 технически особенности)
  - **Какие графики получены и как их интерпретировать**
- Лабораторные для докладов будут распределены случайно (см. гитхаб)