

Алгоритм Tree-ReWeighted Message Passing для вывода в циклических графических моделях

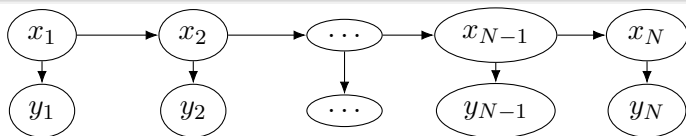
Александр Адуенко

18е мая 2022

Содержание предыдущих лекций

- Выбор априорного распределения. Неинформативные распределения. Распределение Джеффриса.
- EM-алгоритм. Использование EM-алгоритма для отбора признаков в байесовской линейной регрессии.
- Вариационный EM-алгоритм и его использование для вывода в смеси моделей линейной регрессии.
- Гамильтоновы методы Монте-Карло и сравнение с вариационным EM-алгоритмом.
- Ориентированные графические модели и их представление plate notation. Критерий условной независимости d-separation.
- Неориентированные графические модели и их связь с ориентированными.
- Факторные графы и алгоритм Sum-Product для вывода в ациклических графических моделях.
- Скрытые марковские модели (СММ) и алгоритм Витерби. Алгоритм Max-Sum как обобщение алгоритма Витерби.
- Алгоритм Баума-Велча для определения параметров СММ.
- Алгоритмы на основе разрезов графов. Алгоритм α – расширение.

Вывод в графических моделях



$$p(\mathbf{x}, \mathbf{y}) = p(x_1) \prod_{i=2}^N p(x_i | x_{i-1}) \prod_{i=1}^N p(y_i | x_i).$$

Пусть $x_i \in [K]$, $\mathbf{A} = \|a_{ij}\| = \|P(x_l = j | x_{l-1} = i)\|$, $\pi_k = P(x_1 = k)$.

$$p(\mathbf{x}, \mathbf{y} | \mathbf{A}, \boldsymbol{\pi}, \mathbf{B}) = p(x_1 | \boldsymbol{\pi}) \prod_{i=2}^N p(x_i | x_{i-1}, \mathbf{A}) \prod_{i=1}^N p(y_i | x_i, \mathbf{B}).$$

Задачи:

- $p(x_i | \mathbf{y}, \Theta)$ – алгоритм Sum-Product;
- $p(\mathbf{x}_C | \mathbf{y}, \Theta)$ – алгоритм Sum-Product;
- $p(\mathbf{x} | \mathbf{y}, \Theta) \rightarrow \max_{\mathbf{x}}$ – алгоритм Витерби / Max-Sum / Graph-Cut / α – расширение;
- $p(\mathbf{x} | \mathbf{y}, \Theta)$ – сэмплирование;
- $p(\mathbf{y} | \Theta) \rightarrow \max_{\Theta}$ – алгоритм Баума-Велча.

Алгоритм α – расширение

$$\tilde{E}(\mathbf{x}) = \theta_0 + \sum_{(i,j) \in \mathcal{E}} \theta_{ij}(x_i, x_j) + \sum_i \theta_i(x_i) \rightarrow \min_{\mathbf{x}}, x_i \in [K], K \geq 3.$$

Замечание: Задача NP-трудна даже для $K = 3$ и парных потенциалов Поттса.

Идея (α – расширение): Пошагово решать задачи с бинарными переменными.

- 1 Выбираем начальное приближение \mathbf{x} , $x_i \in [K]$;
- 2 В цикле для каждой метки $\alpha \in [K]$ заменяем часть меток на данную, минимизируя энергию;
- 3 Останавливаемся, когда нет улучшений ни для одной метки.

Шаг 2 соответствует введению переменных $q_j \in \{0, 1\}$, что

- $q_j = 0$, если $x_j^{\text{old}} = x_j^{\text{new}}$;
- $q_j = 1$, если $x_j^{\text{old}} \neq \alpha$, $x_j^{\text{new}} = \alpha$.

Иллюстрация работы алгоритма α – расширение



Исходные изображения

Результат сшивки

Иллюстрация работы алгоритма α – расширение 2



Исходные изображения

Результат шивки

Постановка задачи

$$\tilde{E}(\mathbf{x}) = \theta_0 + \sum_{(i,j) \in \varepsilon} \theta_{ij}(x_i, x_j) + \sum_i \theta_i(x_i) \rightarrow \min_{\mathbf{x}}, x_i \in [K], K \geq 3.$$

Замечание: Задача NP-трудна в общем случае даже для $K = 3$ и парных потенциалов Поттса.

- Ациклическая ГМ \implies Точное решение через MaxSum;
- $K = 2$ и субмодулярные потенциалы \implies Точное решение через GraphCut;
- $K \geq 3$ и неравенство треугольника \implies Приближенное решение через α – расширение / $\alpha - \beta$ – замену.

Вопрос 1: Что делать, если $K \geq 3$ и неравенство треугольника не выполнено?

Вопрос 2: Как обработать потенциалы более высоких порядков, например, $\theta(x_1, x_2, x_3)$?

$$\tilde{E}(\mathbf{x}) = \theta_0 + \sum_{(i,j) \in \varepsilon} \theta_{ij}(x_i, x_j) + \sum_i \theta_i(x_i) \rightarrow \min_{\mathbf{x}}, x_i \in [K], K \geq 3.$$

Вопрос: Как обработать $\theta(x_1, x_2, x_3)$?

Идея: Введем $X = x_1 + K(x_2 - 1) + K^2(x_3 - 1) \in [K^3]$.

Тогда $x_1 = \phi_1(X) = X \% K$, $x_2 = \phi_2(X) = 1 + (X \% K^2) / K$,

$x_3 = \phi_3(X) = 1 + X / K^2$,

$\theta(x_1, x_2, x_3) = \tilde{\theta}(X) I[x_1 = \phi_1(X)] I[x_2 = \phi_2(X)] I[x_3 = \phi_3(X)]$.

Алгоритм TRW: LP-релаксация

$$\tilde{E}(\mathbf{x}) = \theta_0 + \sum_{(i,j) \in \varepsilon} \theta_{ij}(x_i, x_j) + \sum_i \theta_i(x_i) \rightarrow \min_{\mathbf{x}}, x_i \in [K], K \geq 3.$$

Введем $z_{ip} = 1 \iff x_i = p$, $z_{ij,pq} = 1 \iff x_i = p, x_j = q$.

Обозначим $\theta_{ip} = \theta_i(p)$, $\theta_{ij}(p, q) = \theta_{ij,pq}$.

$$E(\mathbf{z}, \Theta) = \sum_i \sum_{p=1}^K \theta_{ip} z_{ip} + \sum_{(i,j) \in \varepsilon} \sum_{p,q=1}^K \theta_{ij,pq} z_{ij,pq} \rightarrow \min_{\mathbf{z}}$$

$$\text{s.t. } \sum_p z_{ip} = 1 \forall i, \sum_q z_{ij,pq} = z_{ip}, \sum_p z_{ij,pq} = z_{jq}, z_{ip}, z_{ij,pq} \in \{0, 1\}.$$

LP-релаксация ($\mathbf{z} \in \mathcal{B} \rightarrow \mathbf{z} \in \mathcal{R}$):

$$E(\mathbf{z}, \Theta) = \sum_i \sum_{p=1}^K \theta_{ip} z_{ip} + \sum_{(i,j) \in \varepsilon} \sum_{p,q=1}^K \theta_{ij,pq} z_{ij,pq} \rightarrow \min_{\mathbf{z}}$$

$$\text{s.t. } \sum_p z_{ip} = 1 \forall i, \sum_q z_{ij,pq} = z_{ip}, \sum_p z_{ij,pq} = z_{jq}, z_{ip}, z_{ij,pq} \in [0, 1].$$

$$\min_{\mathbf{z} \in \mathcal{B}} E(\mathbf{z}, \Theta) \geq \min_{\mathbf{z} \in \mathcal{R}} E(\mathbf{z}, \Theta).$$

Алгоритм TRW: Двойственная задача

$$E(\mathbf{z}, \Theta) = \sum_{i \in V} \sum_{p=1}^K \theta_{ip} z_{ip} + \sum_{(i,j) \in \varepsilon} \sum_{p,q=1}^K \theta_{ij,pq} z_{ij,pq} \rightarrow \min_{\mathbf{z}}$$

s.t. $\sum_p z_{ip} = 1 \forall i, \sum_q z_{ij,pq} = z_{ip}, \sum_p z_{ij,pq} = z_{jq}, z_{ip}, z_{ij,pq} \in [0, 1].$

Вопрос: В какой точке достигается минимум в задаче линейного программирования?

Идея: Покроем исходный граф $G = (V, \varepsilon)$ деревьями $G = \cup_{t=1}^T D_t$.

Пусть $n_i \geq 1, n_{ij} \geq 1$ – количество деревьев, в которые входят вершина i и ребро (i, j) соответственно.

Введем $\theta_{ip}^t = \frac{\theta_{ip}}{n_i} I[i \in D_t], \theta_{ij,pq}^t = \frac{\theta_{ij,pq}}{n_{ij}} I[(i, j) \in D_t].$

Тогда $E(\mathbf{z}, \Theta) = \sum_{t=1}^T E^t(\mathbf{z}, \Theta_t), \Theta = \sum_{t=1}^T \Theta_t.$

Задача: $E(\mathbf{z}, \Theta) = \sum_{t=1}^T E^t(\mathbf{z}, \Theta_t) \rightarrow \min_{\mathbf{z} \in \mathcal{R}}.$

Алгоритм TRW: Двойственная задача 2

$$E(\mathbf{z}, \Theta) = \sum_{t=1}^T E^t(\mathbf{z}, \Theta_t) = \sum_{i \in V} \sum_{p=1}^K \theta_{ip} z_{ip} + \sum_{(i,j) \in \varepsilon} \sum_{p,q=1}^K \theta_{ij,pq} z_{ij,pq} \rightarrow \min_{\mathbf{z} \in \mathcal{R}}$$

Идея: Введем \mathbf{z}^t и заменим исходную задачу на эквивалентную

$$\sum_{t=1}^T E^t(\mathbf{z}^t, \Theta_t) \rightarrow \min_{\mathbf{z}_1, \dots, \mathbf{z}_T} \text{ s.t. } \mathbf{z}_t \in \mathcal{R} \forall t, \mathbf{z}_t = \mathbf{z}_1 \forall t \geq 2.$$

$$L(\mathbf{Z}, \Theta, \Lambda) = \sum_{t=1}^T E^t(\mathbf{z}^t, \Theta_t) + \sum_i \sum_{p=1}^K \sum_{t=2}^T \lambda_{ip}^t (z_{ip}^t - z_{ip}^1) +$$

$$\sum_{(i,j) \in \varepsilon} \sum_{p,q=1}^K \sum_{t=2}^T \lambda_{ij,pq}^t (z_{ij,pq}^t - z_{ij,pq}^1) =$$

$$\sum_{t=1}^T \left[\sum_{i \in V} \sum_{p=1}^K (\theta_{ip}^t + \lambda_{ip}^t) z_{ip}^t + \sum_{(i,j) \in \varepsilon} \sum_{p,q=1}^K (\theta_{ij,pq}^t + \lambda_{ij,pq}^t) z_{ij,pq}^t \right],$$

$$\text{где } \Lambda \in \mathcal{L} = \left\{ \Lambda : \lambda_{ip}^1 = - \sum_{t=2}^T \lambda_{ip}^t, \lambda_{ij,pq}^1 = - \sum_{t=2}^T \lambda_{ij,pq}^t \right\}.$$

Алгоритм TRW: Двойственная задача 3

$$E(\mathbf{z}, \Theta) = \sum_{t=1}^T E^t(\mathbf{z}, \Theta_t) = \sum_{i \in V} \sum_{p=1}^K \theta_{ip} z_{ip} + \sum_{(i,j) \in \varepsilon} \sum_{p,q=1}^K \theta_{ij,pq} z_{ij,pq} \rightarrow \min_{\mathbf{z} \in \mathcal{R}}$$

$$L(\mathbf{Z}, \Theta, \Lambda) = \sum_{t=1}^T \left[\sum_{i \in V} \sum_{p=1}^K (\theta_{ip}^t + \lambda_{ip}^t) z_{ip}^t + \sum_{(i,j) \in \varepsilon} \sum_{p,q=1}^K (\theta_{ij,pq}^t + \lambda_{ij,pq}^t) z_{ij,pq}^t \right]$$

$$L(\mathbf{Z}, \Theta, \Lambda) = \sum_{t=1}^T E^t(\mathbf{z}^t, \Theta^t + \Lambda^t), \mathbf{z}^t \in \mathcal{R}, \Lambda^t \in \mathcal{L}.$$

$$\min_{\mathbf{z} \in \mathcal{R}} E(\mathbf{z}, \Theta) \geq \max_{\Lambda \in \mathcal{L}} \min_{\mathbf{z}_1, \dots, \mathbf{z}_T \in \mathcal{R}} L(\mathbf{Z}, \Theta, \Lambda) = \max_{\Lambda \in \mathcal{L}} \sum_{t=1}^T \min_{\mathbf{z}^t \in \mathcal{R}} E^t(\mathbf{z}^t, \Theta^t + \Lambda^t).$$

$$\min_{\mathbf{z} \in \mathcal{R}} E(\mathbf{z}, \Theta) \geq \max_{\Lambda \in \mathcal{L}} \sum_{t=1}^T \min_{\mathbf{z}^t \in \mathcal{R}} E^t(\mathbf{z}^t, \Theta^t + \Lambda^t) = \max_{\Lambda \in \mathcal{L}} \sum_{t=1}^T \min_{\mathbf{z}^t \in \mathcal{B}} E^t(\mathbf{z}^t, \Theta^t + \Lambda^t).$$

Вопрос 1: Что было использовано при получении последнего равенства?

Вопрос 2: Как решить $\min_{\mathbf{z}^t \in \mathcal{B}} E^t(\mathbf{z}^t, \Theta^t + \Lambda^t)$ для фиксированного Λ^t ?

Алгоритм TRW: Двойственная задача 4

$$E(\mathbf{z}, \Theta) = \sum_{t=1}^T E^t(\mathbf{z}, \Theta_t) = \sum_{i \in V} \sum_{p=1}^K \theta_{ip} z_{ip} + \sum_{(i,j) \in \varepsilon} \sum_{p,q=1}^K \theta_{ij,pq} z_{ij,pq} \rightarrow \min_{\mathbf{z} \in \mathcal{R}}$$

Эквивалентная задача:

$$\tilde{E}(\mathbf{Z}, \Theta) = \sum_{t=1}^T E^t(\mathbf{z}^t, \Theta_t) \rightarrow \min_{\mathbf{Z} \in Q}, Q = \{\mathbf{Z} : \mathbf{z}_t \in \mathcal{R} \forall t, \mathbf{z}_t = \mathbf{z}_1 \forall t \geq 2\}.$$

$$\min_{\mathbf{z} \in \mathcal{R}} E(\mathbf{z}, \Theta) = \min_{\mathbf{Z} \in Q} \tilde{E}(\mathbf{Z}, \Theta) = \max_{\Lambda \in \mathcal{L}} \sum_{t=1}^T \min_{\mathbf{z}^t \in \mathcal{B}} E^t(\mathbf{z}^t, \Theta^t + \Lambda^t).$$

Вопрос 1: Что было использовано для замены неравенства на равенство?

$$g(\Lambda) = \sum_{t=1}^T \min_{\mathbf{z}^t \in \mathcal{B}} E^t(\mathbf{z}^t, \Theta^t + \Lambda^t) - \text{вогнутая ли по } \Lambda \text{ на выпуклом } \Lambda \in \mathcal{L}?$$

Hint: $\min_{\mathbf{z}^t \in \mathcal{B}} E^t(\mathbf{z}^t, \Theta^t + \Lambda^t)$ – минимум конечного числа линейных функций по Λ^t , соответствующих разным значениям \mathbf{z}^t .

Вопрос 2: Сколько локальных максимумов имеет $g(\Lambda)$?

Вопрос 3: Дифференцируема ли $g(\Lambda)$?

Вопрос 4: Какой метод стоит использовать для ее оптимизации?

Алгоритм TRW: Двойственная задача 5

$$E^t(\mathbf{z}^t, \Theta^t + \Lambda^t) = \sum_{i \in V} \sum_{p=1}^K (\theta_{ip}^t + \lambda_{ip}^t) z_{ip}^t + \sum_{(i,j) \in \varepsilon} \sum_{p,q=1}^K (\theta_{ij,pq}^t + \lambda_{ij,pq}^t) z_{ij,pq}^t.$$

$$g(\Lambda) = \sum_{t=1}^T \min_{\mathbf{z}^t \in \mathcal{B}} E^t(\mathbf{z}^t, \Theta^t + \Lambda^t) \rightarrow \max_{\Lambda \in \mathcal{L}}.$$

Идея: Использовать метод условного субградиентного подъема по Λ и обновлять $\mathbf{Z}(\Lambda)$ до сходимости.

$$\lambda_{ip}^{t,n} = \lambda_{ip}^{t,n-1} + \alpha_n \left(z_{ip}^{t,n-1} - \frac{\sum_{s: i \in D_s} z_{ip}^{s,n-1}}{n_i} \right);$$

$$\lambda_{ij,pq}^{t,n} = \lambda_{ij,pq}^{t,n-1} + \alpha_n \left(z_{ij,pq}^{t,n-1} - \frac{\sum_{s: (i,j) \in D_s} z_{ij,pq}^{s,n-1}}{n_{ij}} \right);$$

$$\mathbf{z}^{t,n} = \arg \min_{\mathbf{z} \in \mathcal{B}} E^t(\mathbf{z}, \Theta^t + \Lambda^{t,n}).$$

Вопрос 1: Можно ли по $\mathbf{z}^{t,n}$ использовать градиентный шаг?

Вопрос 2: Как в формулах для Λ^n учтено $\Lambda \in \mathcal{L}$?

$$\min_{\mathbf{z} \in \mathcal{B}} E(\mathbf{z}, \theta) \geq \min_{\mathbf{z} \in \mathcal{R}} E(\mathbf{z}, \theta) = \max_{\Lambda \in \mathcal{L}} \sum_{t=1}^T \min_{\mathbf{z}^t \in \mathcal{B}} E^t(\mathbf{z}^t, \Theta^t + \Lambda^t).$$

Вопрос 1: Зависит ли найденная нижняя оценка значения энергии

$$\max_{\Lambda \in \mathcal{L}} \sum_{t=1}^T \min_{\mathbf{z}^t \in \mathcal{B}} E^t(\mathbf{z}^t, \Theta^t + \Lambda^t)$$
 от покрытия графа деревьями?

Вопрос 2: Всегда ли можно покрыть граф $G = (V, \varepsilon)$ деревьями? Как выбрать покрытие деревьями?

Вопрос 3: Как получить приближенное решение \mathbf{z} для задачи

$$\min_{\mathbf{z} \in \mathcal{B}} E(\mathbf{z}, \theta)$$
 после нахождения $\mathbf{z}^t, \Lambda^t, t \in [T]$?

Идея: Рассмотреть ту часть \mathbf{z}^t , где оптимальные значения сходятся. Как согласовать остальные?

- 1 Wainwright, M. J., Jaakkola, T. S., & Willsky, A. S. (2005). MAP estimation via agreement on trees: message-passing and linear programming. *IEEE transactions on information theory*, 51(11), 3697-3717.
- 2 Kolmogorov, V. (2005, January). Convergent tree-reweighted message passing for energy minimization. In *International Workshop on Artificial Intelligence and Statistics* (pp. 182-189). PMLR.
- 3 Kolmogorov, V., & Wainwright, M. (2012). On the optimality of tree-reweighted max-product message-passing. *arXiv preprint arXiv:1207.1395*.
- 4 Kolmogorov, V. (2014). A new look at reweighted message passing. *IEEE transactions on pattern analysis and machine intelligence*, 37(5), 919-930.
- 5 Koller, Daphne, and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.