Вычислительные методы приближенного подсчета нормировочной константы марковского случайного поля

Александр Новиков

Научный руководитель:
Дмитрий Ветров

25 апреля 2015 г.

Обзор

- Марковское случайное поле – мощный инструмент работы со структурированными данными.
- Его применение ограничивать сложность расчёта нормировочной константы.
- Мы предлагаем три новых методы приближения норм. константы:
  - обобщение метода n-окрестностей (Крыжановский & Литинский, 2014);
  - метод, основанный на разложении в ряд Тейлора;
  - метод, основанный на разложении Tensor Train (Оселедец, 2011).

Марковское случайное поле

Марковское случайное поле (Markov random field, MRF) – способ компактно задать вероятностное распределение:

$$P(\boldsymbol{x}) = \frac{1}{Z}\prod_{\ell=1}^{m} \Psi_\ell(\boldsymbol{x})$$

где Ψ_ℓ зависит от небольшого подмножества переменных.

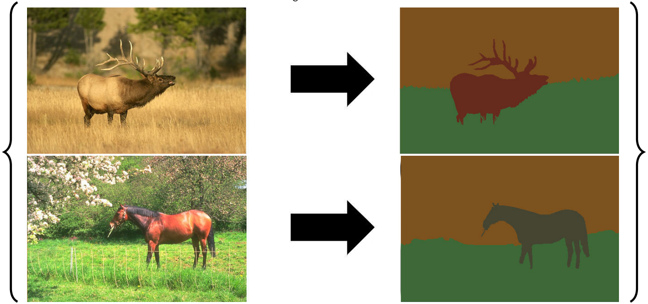
Обозначения:

- $\boldsymbol{x} = (x_1, \ldots, x_n)$ – переменные ($x_i \in \{1, \ldots, d_i\}$);
- $\widehat{P}(\boldsymbol{x}) = \prod_{\ell=1}^{m} \Psi_\ell(\boldsymbol{x})$ – ненормированное распределение Гиббса;
- $Z = \sum_{\boldsymbol{x}} \widehat{P}(\boldsymbol{x})$ – нормировочная константа;
- $E(\boldsymbol{x}) = -\ln \widehat{P}(\boldsymbol{x})$ – энергия.

Прикладная задача

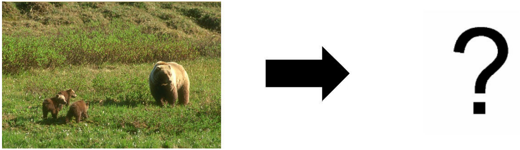
Обучение

X_{train}



T_{train}

Контроль



Обучение MRF

Модель марковского случайного поля:

$$P(T|X,W) = \frac{1}{Z(X,W)} \prod_{\ell=1}^{m} \Psi_\ell(T_\ell; X, W)$$

Оценка параметров по методу максимального правдоподобия:

$$\begin{aligned} W^* &= \arg\max_W P(T_{\text{train}}|X_{\text{train}}, W) = \\ &= \arg\max_W \frac{1}{Z(X_{\text{train}, W})} \prod_\ell \Psi_\ell(T_{\text{train}}; X_{\text{train}}, W) = \\ &= \arg\max_W \left(\sum_\ell \ln \Psi_\ell(T_{\text{train}}; X_{\text{train}}, W) - \ln Z(X_{\text{train}}, W) \right) \end{aligned}$$

Метод n-окрестностей

1. Новый метод (2014) для оценки Z из статистической физики;
2. Предположение метода: подмножества («слои») значений энергий MRF можно считать выборками из нормального распределение;
3. Требует численного взятия одномерного интеграла.

Недостатки

1. Работает только для модели Изинга;
2. Точность (предположение нарушается на практике);
3. Скорость работы ограничивает применение метода к обучению MRF.

Метод «суммировать всё»

Пусть эмпирическая ф.р. множества всех энергий близка к нормальному распределению:

$$P(E < y) = \frac{\sum_{\boldsymbol{x}} [E(\boldsymbol{x}) < y]}{|\boldsymbol{x}|} \approx \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(t-\mu)^2}{2\sigma^2}) dt$$

Тогда нормировочная константа приближается как:

$$\ln Z \approx \ln \left(|\boldsymbol{x}| \int_{-\infty}^{\infty} f(E \mid \mu, \sigma^2) \exp(-E) dE \right) = -\mu + \frac{\sigma^2}{2} + \ln(|\boldsymbol{x}|)$$

Нами получены аналитические формулы для μ и σ.

1. Работает для любых MRF;
2. Очень быстрый.

Ряд Тейлора

Рассмотрим определение нормировочной константы:

$$Z = \sum_{\boldsymbol{x}} \exp(-E(\boldsymbol{x}))$$

Разложим показательную функцию в ряд Тейлора:

$$Z \approx \sum_{\boldsymbol{x}} \left(\exp(-E_0) - \exp(-E_0)(E(\boldsymbol{x}) - E_0) + \exp(-E_0)\frac{1}{2}(E(\boldsymbol{x}) - E_0)^2 \right)$$

$$\ln Z = -E_0 + \ln\left(1 - \mu(1 + E_0) + \frac{1}{2}\mu^2 + \frac{1}{2}\sigma^2 + E_0 + \frac{1}{2}E_0^2\right) + \ln(|\boldsymbol{x}|)$$

1. Работает для любых MRF;
2. Очень быстрый;
3. Близок к «суммировать всё» при $E_0 = \mu$.

Tensor Train

Выведено точное выражение для Z через произведение экспоненциально больших матриц:

$$Z = \underbrace{A_1}_{1\times 2^m} \cdots \underbrace{A_n}_{2^m \times 1},$$

Нами получены аналитические выражения для представления матриц A_i в формате Tensor Train (TT).

Пользуясь приближенным умножением TT-матриц можно оценить Z.

1. Работает для любых MRF;
2. Контролируемая точность;
3. Относительно медленный.

Теоретический анализ

Пользуясь следующей теоремой, можно по требуемой точности расчета нормировочной константы выбрать параметр точности умножения ТТ-матриц.

Теорема

Ошибка алгоритма оценки нормировочной константы через ТТ-разложение ограничена следующим образом:

$$\left| Z - \widehat{Z} \right| \leq \|A_1\|_2 \cdots \|A_n\|_2 \left((1 + \varepsilon)^{n-1} - 1 \right),$$

где ε — абсолютная точность умножения ТТ-матриц.

Эксперименты

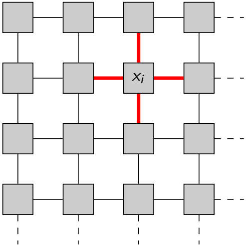
Рассмотрим модель Изинга:

$$P(\boldsymbol{x}) = \frac{1}{Z}\prod_{i=1}^{n}\exp\left(-\frac{1}{T}h_i x_i\right)\prod_{(i,j)\in\mathcal{E}}\exp\left(-\frac{1}{T}c_{ij}x_i x_j\right)$$

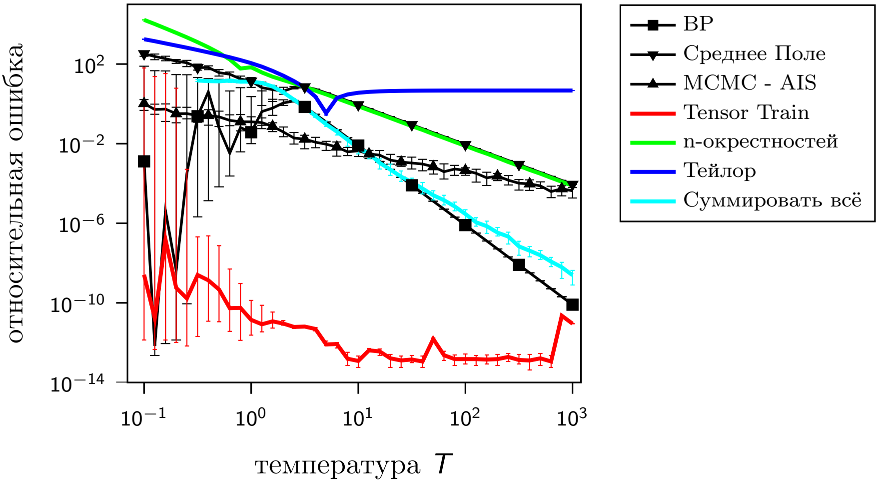
где $x_i \in \{-1, 1\}$.

Обозначения:

- Температура T;
- Унарные коэффициента h_i;
- Парные коэффициенты c_{ij}.



Эксперимент №1



Модель Изинга размера 10×10.

Выводы

1. Если нужна высокая точность работы, TT работает лучше всех;
2. Если нужна быстрая и не очень точная оценка, Тейлор и суммировать всё интересные новые методы, применимые к широкому классу MRF;
3. По результатам работы сделаны доклады на ведущих мировых конференциях и опубликована статья в журнале ВАК.

На защиту выносится

1. Метод «суммировать всё» для быстрой оценки нормировочной константы MRF;
2. Метод, основанный на разложении в ряд Тейлора;
3. Тензорный метод аппроксимации нормировочной константы на основе разложения Tensor Train;
4. Теоретические оценки точности тензорного метода.