Сажаем MRF на тензорный поезд

Новиков А. В.,
Родоманов А. О., Осокин А. А., Ветров Д. П.

МГУ, ВМиК, каф. ММП

Спецсеминар «Байесовские методы машинного обучения»

Оглавление

Обозначения

Опр. *Тензор* — многомерный массив (функция индексов)

$$\mathbf{A}(x) = \mathbf{A}(x_1, \ldots, x_n), \quad x_i \in \{1, \ldots, d_i\}$$

Терминология:

- n — *размерность* тензора;
- x_i — *индексы*.

Будем использовать

- большие жирные буквы ($\mathbf{A}$) для обозначения тензоров;
- большие буквы (A) для обозначения матриц;
- маленькие жирные буквы ($\boldsymbol{x}$) для обозначения векторов.

TT-представление

Опр. Тензор $\mathbf{A}$ представлен в *TT-формате*, если

$$\mathbf{A}(x_1, \ldots, x_n) = G_1^{\mathbf{A}}[x_1] G_2^{\mathbf{A}}[x_2] \ldots G_n^{\mathbf{A}}[x_n],$$

где $G_i^{\mathbf{A}}[x_i]$ — матрица размера $\mathrm{r}_{i-1}(\mathbf{A}) \times \mathrm{r}_i(\mathbf{A})$, $\mathrm{r}_0(\mathbf{A}) = \mathrm{r}_n(\mathbf{A}) = 1$.

Терминология:

- $G_i^{\mathbf{A}}$ — *TT-ядра*;
- $\mathrm{r}_i(\mathbf{A})$ — *TT-ранги*;
- $r(\mathbf{A}) = \max\limits_{i=0,\ldots,n} r_i(\mathbf{A})$ — *максимальный TT-ранг*.

Замечание: TT-формат требует $O\left(ndr^2(\mathbf{A})\right)$ памяти для хранения $O\left(d^n\right)$ элементов ($d = \max\limits_{i=1,\ldots,n} d_i$).

Пример

$$\mathbf{A}(x_1, x_2, x_3) = x_1 + x_2 + x_3,$$

$$x_1 \in \{1, 2, 3\},\ x_2 \in \{1, 2, 3, 4\},\ x_3 \in \{1, 2, 3, 4, 5\}.$$

$$\mathbf{A}(x_1, x_2, x_3) = G_1^{\mathbf{A}}[x_1] G_2^{\mathbf{A}}[x_2] G_3^{\mathbf{A}}[x_3],$$

Пример

$$\mathbf{A}(x_1, x_2, x_3) = x_1 + x_2 + x_3,$$

$$x_1 \in \{1, 2, 3\},\ x_2 \in \{1, 2, 3, 4\},\ x_3 \in \{1, 2, 3, 4, 5\}.$$

$$\mathbf{A}(x_1, x_2, x_3) = G_1^{\mathbf{A}}[x_1] G_2^{\mathbf{A}}[x_2] G_3^{\mathbf{A}}[x_3],$$

$$G_1^{\mathbf{A}}[x_1] = \begin{bmatrix} x_1 & 1 \end{bmatrix} \quad G_2^{\mathbf{A}}[x_2] = \begin{bmatrix} 1 & 0 \\ x_2 & 1 \end{bmatrix} \quad G_3^{\mathbf{A}}[x_3] = \begin{bmatrix} 1 \\ x_3 \end{bmatrix}$$

Проверим:

$$A(x_1, x_2, x_3) = \begin{bmatrix} x_1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ x_2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ x_3 \end{bmatrix} =$$

$$= \begin{bmatrix} x_1 + x_2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ x_3 \end{bmatrix} = x_1 + x_2 + x_3.$$

Пример

$$\mathbf{A}(x_1, x_2, x_3) = x_1 + x_2 + x_3,$$
$$x_1 \in \{1, 2, 3\},\ x_2 \in \{1, 2, 3, 4\},\ x_3 \in \{1, 2, 3, 4, 5\}.$$

$$\mathbf{A}(x_1, x_2, x_3) = G_1^{\mathbf{A}}[x_1] G_2^{\mathbf{A}}[x_2] G_3^{\mathbf{A}}[x_3],$$

$$G_1^{\mathbf{A}}[x_1] = \begin{bmatrix} x_1 & 1 \end{bmatrix} \quad G_2^{\mathbf{A}}[x_2] = \begin{bmatrix} 1 & 0 \\ x_2 & 1 \end{bmatrix} \quad G_3^{\mathbf{A}}[x_3] = \begin{bmatrix} 1 \\ x_3 \end{bmatrix}$$

$$G_1^{\mathbf{A}} = \left(\begin{bmatrix} 1 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \end{bmatrix}, \begin{bmatrix} 3 & 1 \end{bmatrix} \right)$$
$$G_2^{\mathbf{A}} = \left(\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \right)$$
$$G_3^{\mathbf{A}} = \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \end{bmatrix} \right)$$

Количество элементов в тензоре: $3 \cdot 4 \cdot 5 = 60$.
TT-формат требует хранения 32-х чисел.

Сложение тензоров

Даны $\mathbf{A}$ и $\mathbf{B}$ в TT-формате:

$$\mathbf{A}(x) = G_1^{\mathbf{A}}[x_1] \dots G_n^{\mathbf{A}}[x_n], \quad \mathbf{B}(x) = G_1^{\mathbf{B}}[x_1] \dots G_n^{\mathbf{B}}[x_n].$$

Найти TT-представление

$$\mathbf{C} = \mathbf{A} + \mathbf{B},$$

$$\mathbf{C}(x) = \mathbf{A}(x) + \mathbf{B}(x).$$

Сложение тензоров

Даны $\mathbf{A}$ и $\mathbf{B}$ в TT-формате:

$$\mathbf{A}(x) = G_1^{\mathbf{A}}[x_1] \dots G_n^{\mathbf{A}}[x_n], \quad \mathbf{B}(x) = G_1^{\mathbf{B}}[x_1] \dots G_n^{\mathbf{B}}[x_n].$$

Найти TT-представление

$$\mathbf{C} = \mathbf{A} + \mathbf{B},$$

$$\mathbf{C}(x) = \mathbf{A}(x) + \mathbf{B}(x).$$

TT-ядра определяются следующим образом:

$$G_i^{\mathbf{C}}[x_i] = \begin{bmatrix} G_i^{\mathbf{A}}[x_i] & 0 \\ 0 & G_i^{\mathbf{B}}[x_i] \end{bmatrix}, \quad i = 2, \dots, n-1,$$

$$G_1^{\mathbf{C}}[x_1] = \begin{bmatrix} G_1^{\mathbf{A}}[x_1] & G_1^{\mathbf{B}}[x_1] \end{bmatrix}, \quad G_n^{\mathbf{C}}[x_n] = \begin{bmatrix} G_n^{\mathbf{A}}[x_n] \\ G_n^{\mathbf{B}}[x_n] \end{bmatrix}.$$

TT-ранги складываются.

Умножение тензора на число

Пусть $\mathbf{A}(x) = G_1^{\mathbf{A}}[x_1] G_2^{\mathbf{A}}[x_2] \dots G_n^{\mathbf{A}}[x_n], \quad c \in \mathbb{R}.$

Найти TT-представление

$$\mathbf{B} = c \cdot \mathbf{A},$$

$$\mathbf{B}(x) = c \cdot \mathbf{A}(x).$$

Умножение тензора на число

Пусть $\mathbf{A}(x) = G_1^{\mathbf{A}}[x_1] G_2^{\mathbf{A}}[x_2] \dots G_n^{\mathbf{A}}[x_n], \quad c \in \mathbb{R}.$

Найти TT-представление

$$\mathbf{B} = c \cdot \mathbf{A},$$

$$\mathbf{B}(x) = c \cdot \mathbf{A}(x).$$

Одно из TT-ядер умножается на c, TT-ранги не изменяются.

TT-округление

Пусть имеется TT-представление $\mathbf{A}(x) = G_1^{\mathbf{A}}[x_1] \dots G_n^{\mathbf{A}}[x_n]$, ранги которого неоптимальные.

Процедура TT-округления [Oseledets, 2011]

$$\widehat{\mathbf{A}} = \text{round}(\mathbf{A}, \varepsilon), \quad \varepsilon \geq 0$$

находит тензор $\widehat{\mathbf{A}}$:

1. $\|\mathbf{A} - \widehat{\mathbf{A}}\|_F \leq \varepsilon \|\mathbf{A}\|_F$;
2. его TT-ранги минимальны среди всех тензоров $\mathbf{B}$: $\|\mathbf{A} - \mathbf{B}\|_F \leq \frac{\varepsilon}{\sqrt{n-1}} \|\mathbf{A}\|_F$.

Где $\|\mathbf{A}\|_F = \sqrt{\sum_{x_1, \dots, x_n} \mathbf{A}^2(x_1, \dots, x_n)}$.

Операции над тензорами в TT-формате

Операция	Ранг результата
$\mathbf{C} = c \cdot \mathbf{A}$	$\mathsf{r}(\mathbf{C}) = \mathsf{r}(\mathbf{A})$
$\mathbf{C} = \mathbf{A} + c$	$\mathsf{r}(\mathbf{C}) = \mathsf{r}(\mathbf{A}) + 1$
$\mathbf{C} = \mathbf{A} + \mathbf{B}$	$\mathsf{r}(\mathbf{C}) \leq \mathsf{r}(\mathbf{A}) + \mathsf{r}(\mathbf{B})$
$\mathbf{C} = \mathbf{A} \odot \mathbf{B}$	$\mathsf{r}(\mathbf{C}) \leq \mathsf{r}(\mathbf{A})\,\mathsf{r}(\mathbf{B})$
$\mathbf{C} = \mathsf{round}(\mathbf{A}, \varepsilon)$	$\mathsf{r}(\mathbf{C}) \leq \mathsf{r}(\mathbf{A})$
$\mathsf{sum}\,\mathbf{A}$	–
$\|\mathbf{A}\|_F$	–

Как найти TT-представление тензора

- Для некоторых частных случаев существуют аналитические формулы;
  Пример.

$$\mathbf{A}(x_1, \ldots, x_n) = \sum_{i=1}^{n} f_i(x_i),$$

$$G_i^{\mathbf{A}}[x_i] = \begin{bmatrix} 1 & 0 \\ f_i(x_i) & 1 \end{bmatrix}, \quad i = 2, \ldots, n-1,$$

$$G_1^{\mathbf{A}}[x_1] = \begin{bmatrix} f_1(x_1) & 1 \end{bmatrix}, \quad G_n^{\mathbf{A}}[x_n] = \begin{bmatrix} 1 \\ f_n(x_n) \end{bmatrix}.$$

Как найти TT-представление тензора

- Для некоторых частных случаев существуют аналитические формулы;
- Для небольших тензоров есть точный алгоритм TT-SVD [Oseledets, 2011]. Например, для тензора с числом элементов $5^8 \approx 400\,000$ метод работает 0.5 сек. на моём ноутбуке;
- Для больших тензоров есть приближенные алгоритмы, которые строят TT-представление по небольшому подмножеству элементов тензора: DMRG-cross [Savostyanov and Oseledets, 2011] и AMEn-cross [Dolgov and Savostyanov, 2013].

Марковское случайное поле

Обозначения:

- $\boldsymbol{x} = (x_1, \ldots, x_n)$ — переменные ($x_i \in \{1, \ldots, d_i\}$);
- $\boldsymbol{\Theta}_\ell(\boldsymbol{x}) = \boldsymbol{\Theta}_\ell(\boldsymbol{x}^\ell)$ — потенциалы ($\ell = 1, \ldots, m$);
- $\boldsymbol{x}^\ell$ — подмножество переменных, от которых зависит $\boldsymbol{\Theta}_\ell$;
- $\mathbf{E}(\boldsymbol{x}) = \sum_{\ell=1}^{m} \boldsymbol{\Theta}_\ell(\boldsymbol{x}^\ell)$ — энергия;
- $\widehat{\mathbf{P}}(\boldsymbol{x}) = \exp(-\mathbf{E}(\boldsymbol{x}))$ — ненормированное распределение Гиббса;
- $Z = \sum_{\boldsymbol{x}} \widehat{\mathbf{P}}(\boldsymbol{x})$ — нормировочная константа;
- $\boldsymbol{\Psi}_\ell(\boldsymbol{x}^\ell) = \exp(-\boldsymbol{\Theta}_\ell(\boldsymbol{x}^\ell))$ — факторы.

Вероятностная интерпретация

Построение TT-разложения тензора вероятностей $\mathbf{P} = \frac{1}{Z}\widehat{\mathbf{P}}$ имеет смысл добавлений скрытых переменных α_i, $i = 0, \ldots, n$:

$$\mathbf{P}(\boldsymbol{x}) = G_1^{\mathbf{P}}[x_1] G_2^{\mathbf{P}}[x_2] \ldots G_n^{\mathbf{P}}[x_n] =$$
$$= \sum_{\alpha_{\mathbf{0}}, \ldots, \alpha_{\boldsymbol{n}}} G_1^{\mathbf{P}}[x_1](\alpha_0, \alpha_1) G_2^{\mathbf{P}}[x_2](\alpha_1, \alpha_2) \ldots G_n^{\mathbf{P}}[x_n](\alpha_{n-1}, \alpha_n).$$

Таким образом,

$$\mathbf{P}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}} \mathbf{P}(\boldsymbol{x}, \boldsymbol{\alpha}),$$

где α_i принимает значения от 1 до $\mathrm{r}_i(\mathbf{P})$.

Вероятностная интерпретация

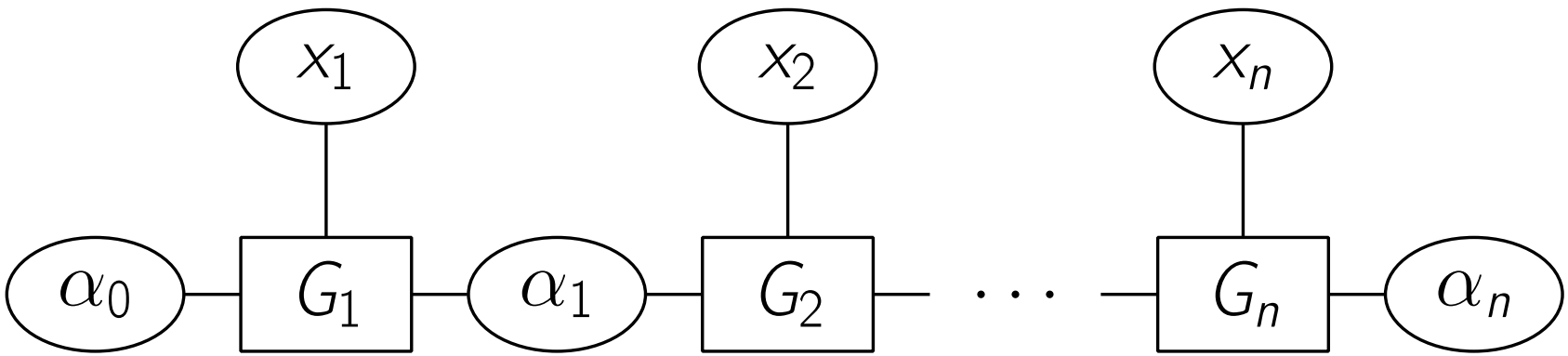
Построение TT-разложения тензора вероятностей $\mathbf{P} = \frac{1}{Z}\widehat{\mathbf{P}}$ имеет смысл добавлений скрытых переменных α_i, $i = 0, \ldots, n$:

$$\mathbf{P}(\boldsymbol{x}) = G_1^{\mathbf{P}}[x_1] G_2^{\mathbf{P}}[x_2] \ldots G_n^{\mathbf{P}}[x_n] =$$
$$= \sum_{\alpha_0, \ldots, \alpha_n} G_1^{\mathbf{P}}[x_1](\alpha_0, \alpha_1) G_2^{\mathbf{P}}[x_2](\alpha_1, \alpha_2) \ldots G_n^{\mathbf{P}}[x_n](\alpha_{n-1}, \alpha_n).$$

Таким образом,

$$\mathbf{P}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}} \mathbf{P}(\boldsymbol{x}, \boldsymbol{\alpha}),$$

где α_i принимает значения от 1 до $\mathrm{r}_i(\mathbf{P})$.



Алгоритм преобразования энергии в TT-формат

Алгоритм:

1. Представить каждый отдельный потенциал $\boldsymbol{\Theta}_\ell(\boldsymbol{x}^\ell)$ в TT-формате;
2. По TT-представлению $\boldsymbol{\Theta}_\ell(\boldsymbol{x}^\ell)$ получить TT-представление $\boldsymbol{\Theta}_\ell(\boldsymbol{x})$.
3. Найти сумму всех потенциалов: $\mathbf{E} = \sum_{\ell=1}^{m} \boldsymbol{\Theta}_\ell$.

Алгоритм преобразования энергии в TT-формат

Алгоритм:

1. Представить каждый отдельный потенциал $\boldsymbol{\Theta}_\ell(\boldsymbol{x}^\ell)$ в TT-формате;
2. По TT-представлению $\boldsymbol{\Theta}_\ell(\boldsymbol{x}^\ell)$ получить TT-представление $\boldsymbol{\Theta}_\ell(\boldsymbol{x})$.
3. Найти сумму всех потенциалов: $\mathbf{E} = \sum_{\ell=1}^{m} \boldsymbol{\Theta}_\ell$.

Пример. Рассмотрим MRF с переменными x_1, x_2, x_3, x_4, x_5 и потенциал $\boldsymbol{\Theta}_\ell(\boldsymbol{x}^\ell)$, который зависит только от переменных x_1, x_2, x_4 (т. е. $\boldsymbol{x}^\ell = (x_1, x_2, x_4)$).
После первого шага алгоритма получаем:

$$\boldsymbol{\Theta}_\ell(x_1, x_2, x_4) = G_1^\ell[x_1] G_2^\ell[x_2] G_4^\ell[x_4],$$
$$\mathrm{r}_0(\boldsymbol{\Theta}_\ell) = \mathrm{r}_4(\boldsymbol{\Theta}_\ell) = 1,$$
$$\mathrm{r}_2(\boldsymbol{\Theta}_\ell) = \mathrm{r}_3(\boldsymbol{\Theta}_\ell).$$

Определим $G_3^\ell[x_3] \equiv I_{\mathbf{r_2}} = I_{\mathbf{r_3}}$ и $G_5^\ell[x_5] \equiv I_{\mathbf{r_4}} = I_1$.

TT-ранги энергии

Теорема. Если порядок потенциалов не превосходит p, то максимальный TT-ранг тензора энергии $\mathbf{E}$, построенного алгоритмом, ограничен сверху:

$$\mathrm{r}(\mathbf{E}) \leq d^{\frac{p}{2}} m,$$

где

- d — количество значений каждой переменной;
- m — количество потенциалов;
- p — максимальный порядок потенциалов.

TT-ранги энергии

Теорема. Если порядок потенциалов не превосходит p, то максимальный TT-ранг тензора энергии $\mathbf{E}$, построенного алгоритмом, ограничен сверху:

$$\mathrm{r}(\mathbf{E}) \leq d^{\frac{p}{2}} m,$$

где

- d — количество значений каждой переменной;
- m — количество потенциалов;
- p — максимальный порядок потенциалов.

Например, для модели Изинга на квадратной четырехсвязной $(q \times q)$-решетке: $\mathrm{r}(\mathbf{E}) \leq 2 \cdot (3q^2 - 2q)$.

Преобразование вероятности в TT-формат

$\widehat{\mathbf{P}}$ можно представить тем же способом:

$$\widehat{\mathbf{P}} = \bigodot_{\ell=1}^{m} \boldsymbol{\Psi}_\ell.$$

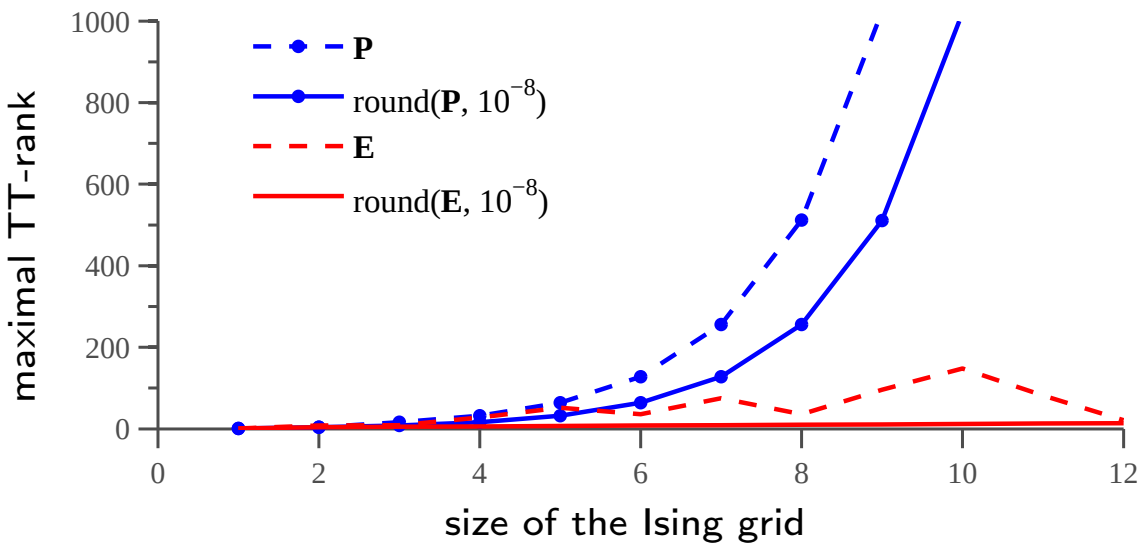
Тем не менее, TT-ранги $\widehat{\mathbf{P}}$ растут экспоненциально.

Преобразование вероятности в TT-формат

$\widehat{\mathbf{P}}$ можно представить тем же способом:

$$\widehat{\mathbf{P}} = \bigodot_{\ell=1}^{m} \boldsymbol{\Psi}_\ell.$$

Тем не менее, TT-ранги $\widehat{\mathbf{P}}$ растут экспоненциально.



Мотивация

1. TT-ранги тензора вероятностей $\widehat{\mathbf{P}}$ экспоненциальные;
2. Нужен алгоритм, который не будет напрямую строить TT-представление $\widehat{\mathbf{P}}$.

Векторы в TT-формате

Пусть задано соответствие между индексами вектора $\boldsymbol{b}$ и n-мерными векторами: $j \leftrightarrow (y_1, \ldots, y_n)$.

Опр. *Вектор $\boldsymbol{b}$ представлен в TT-формате*, если задано TT-представление тензора $\mathbf{B}$:

$$\mathbf{B}(y_1, \ldots, y_n) = \boldsymbol{b}(j).$$

Векторы в TT-формате

Пусть задано соответствие между индексами вектора $\boldsymbol{b}$ и n-мерными векторами: $j \leftrightarrow (y_1, \ldots, y_n)$.

Опр. *Вектор $\boldsymbol{b}$ представлен в TT-формате*, если задано TT-представление тензора $\mathbf{B}$:

$$\mathbf{B}(y_1, \ldots, y_n) = \boldsymbol{b}(j).$$

Пример. Дан вектор $\boldsymbol{b}$ из 18-ти элементов.
Рассмотрим отображение $j \leftrightarrow (y_1, y_2, y_3)$,
где $y_1 \in \{1, 2\}$, $y_2 \in \{1, 2, 3\}$, $y_3 \in \{1, 2, 3\}$:

$$\begin{aligned} 1 &\leftrightarrow (1,1,1) & \mathbf{B}(1,1,1) &= \boldsymbol{b}(1), \\ 2 &\leftrightarrow (1,1,2) & \mathbf{B}(1,1,2) &= \boldsymbol{b}(2), \\ &\ldots & &\ldots \\ 18 &\leftrightarrow (2,3,3) & \mathbf{B}(2,3,3) &= \boldsymbol{b}(18). \end{aligned}$$

Произведение Кронекера

Опр. Пусть A — матрица размеров $p \times q$, а B — матрица размеров $h \times k$. *Произведением Кронекера* матриц A и B называется блочная матрица

$$M = A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1q}B \\ \vdots & \ddots & \vdots \\ a_{p1}B & \dots & a_{pq}B \end{bmatrix}$$

размеров $ph \times qk$.

$$M(x_1, x_2; y_1, y_2) = A(x_1, y_1)B(x_2, y_2).$$

Мотивация TT-формата для матриц

Пусть $M = A \otimes B \otimes C$, то есть

$$M(x_1, x_2, x_3; y_1, y_2, y_3) = A(x_1, y_1)B(x_2, y_2)C(x_3, y_3).$$

Мотивация TT-формата для матриц

Пусть $M = A \otimes B \otimes C$, то есть

$$M(x_1, x_2, x_3;\, y_1, y_2, y_3) = A(x_1, y_1)B(x_2, y_2)C(x_3, y_3).$$

Введем составные индексы: $z_1 = (x_1, y_1), z_2 = (x_2, y_2), z_3 = (x_3, y_3)$.
Тогда

$$\mathbf{M}(z_1, z_2, z_3) = A(z_1)B(z_2)C(z_3).$$

Мотивация TT-формата для матриц

Пусть $M = A \otimes B \otimes C$, то есть

$$M(x_1, x_2, x_3;\ y_1, y_2, y_3) = A(x_1, y_1)B(x_2, y_2)C(x_3, y_3).$$

Введем составные индексы: $z_1 = (x_1, y_1), z_2 = (x_2, y_2), z_3 = (x_3, y_3)$.
Тогда

$$\mathbf{M}(z_1, z_2, z_3) = A(z_1)B(z_2)C(z_3).$$

Обозначим $G_1^{\mathbf{M}}[z_1] = A(z_1)$, $G_2^{\mathbf{M}}[z_2] = B(z_2)$, $G_3^{\mathbf{M}}[z_3] = C(z_3)$.
Получаем TT-представление для $\mathbf{M}$:

$$\mathbf{M}(z_1, z_2, z_3) = G_1^{\mathbf{M}}[z_1]G_2^{\mathbf{M}}[z_2]G_3^{\mathbf{M}}[z_3].$$

Мотивация TT-формата для матриц

Пусть $M = A \otimes B \otimes C$, то есть

$$M(x_1, x_2, x_3;\ y_1, y_2, y_3) = A(x_1, y_1)B(x_2, y_2)C(x_3, y_3).$$

Введем составные индексы: $z_1 = (x_1, y_1), z_2 = (x_2, y_2), z_3 = (x_3, y_3)$.
Тогда

$$\mathbf{M}(z_1, z_2, z_3) = A(z_1)B(z_2)C(z_3).$$

Обозначим $G_1^{\mathbf{M}}[z_1] = A(z_1), G_2^{\mathbf{M}}[z_2] = B(z_2), G_3^{\mathbf{M}}[z_3] = C(z_3)$.
Получаем TT-представление для $\mathbf{M}$:

$$\mathbf{M}(z_1, z_2, z_3) = G_1^{\mathbf{M}}[z_1]G_2^{\mathbf{M}}[z_2]G_3^{\mathbf{M}}[z_3].$$

Таким образом:

$$\mathbf{M}((x_1, y_1), (x_2, y_2), (x_3, y_3)) = G_1^{\mathbf{M}}[(x_1, y_1)]G_2^{\mathbf{M}}[(x_2, y_2)]G_3^{\mathbf{M}}[(x_3, y_3)].$$

Матрицы в TT-формате

Пусть задано соответствие между индексами матрицы $M = [M(i,j)]$ и n-мерными векторами: $i \leftrightarrow (x_1, \ldots, x_n)$ и $j \leftrightarrow (y_1, \ldots, y_n)$.

Опр. *Матрица M представлена в TT-формате*, если задано TT-представление тензора $\mathbf{M}$:

$$\mathbf{M}((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)) = M(i,j),$$

то есть задано представление

$$\mathbf{M}((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)) = G_1^{\mathbf{M}}[(x_1, y_1)] \ldots G_n^{\mathbf{M}}[(x_n, y_n)].$$

Матрицы в TT-формате

Пусть задано соответствие между индексами матрицы $M = [M(i,j)]$ и n-мерными векторами: $i \leftrightarrow (x_1, \ldots, x_n)$ и $j \leftrightarrow (y_1, \ldots, y_n)$.

Опр. *Матрица M представлена в TT-формате*, если задано TT-представление тензора $\mathbf{M}$:

$$\mathbf{M}((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)) = M(i,j),$$

то есть задано представление

$$\mathbf{M}((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)) = G_1^{\mathbf{M}}[(x_1, y_1)] \ldots G_n^{\mathbf{M}}[(x_n, y_n)].$$

Ещё две операции:

Операция	Ранг результата
$M = M_1 M_2$	$\mathrm{r}(M) \leq \mathrm{r}(M_1)\,\mathrm{r}(M_2)$
$\boldsymbol{f} = M\boldsymbol{b}$	$\mathrm{r}(\boldsymbol{f}) \leq \mathrm{r}(M)\,\mathrm{r}(\boldsymbol{b})$

Алгоритм оценки нормировочной константы – 1

Нормировочная константа $Z = \sum\limits_{\boldsymbol{x}} \widehat{\mathbf{P}}(\boldsymbol{x})$.

Преобразуем $\widehat{\mathbf{P}}(\boldsymbol{x})$.

Свойство произведения Кронекера:

$$a \cdot b = a \otimes b, \quad a \in \mathbb{R},\ b \in \mathbb{R}.$$

Применяя данное свойство, получаем:

$$\widehat{\mathbf{P}}(\boldsymbol{x}) = \prod_{\ell=1}^{m} \underbrace{\boldsymbol{\Psi}_\ell(\boldsymbol{x})}_{\in \mathbb{R}} = \bigotimes_{\ell=1}^{m} \boldsymbol{\Psi}_\ell(\boldsymbol{x}) = \bigotimes_{\ell=1}^{m} \left(G_1^\ell[x_1] \dots G_n^\ell[x_n] \right).$$

Алгоритм оценки нормировочной константы – 1

Нормировочная константа $Z = \sum_{\boldsymbol{x}} \widehat{\mathbf{P}}(\boldsymbol{x})$.

Преобразуем $\widehat{\mathbf{P}}(\boldsymbol{x})$.

Свойство произведения Кронекера:

$$a \cdot b = a \otimes b, \quad a \in \mathbb{R},\ b \in \mathbb{R}.$$

Применяя данное свойство, получаем:

$$\widehat{\mathbf{P}}(\boldsymbol{x}) = \prod_{\ell=1}^{m} \underbrace{\boldsymbol{\Psi}_\ell(\boldsymbol{x})}_{\in \mathbb{R}} = \bigotimes_{\ell=1}^{m} \boldsymbol{\Psi}_\ell(\boldsymbol{x}) = \bigotimes_{\ell=1}^{m} \left(G_1^\ell[x_1] \dots G_n^\ell[x_n] \right).$$

Свойство смешанного произведения:

$$AC \otimes BD = (A \otimes B)(C \otimes D).$$

Таким образом:

$$\widehat{\mathbf{P}}(\boldsymbol{x}) = \left(G_1^1[x_1] \otimes \dots \otimes G_1^m[x_1] \right) \dots \left(G_n^1[x_n] \otimes \dots \otimes G_n^m[x_n] \right).$$

Алгоритм оценки нормировочной константы – 2

$$\widehat{\mathbf{P}}(x) = \left(G_1^1[x_1] \otimes \ldots \otimes G_1^m[x_1]\right) \ldots \left(G_n^1[x_n] \otimes \ldots \otimes G_n^m[x_n]\right).$$

Обозначим: $A_i[x_i] = G_i^1[x_i] \otimes \ldots \otimes G_i^m[x_i]$.

Алгоритм оценки нормировочной константы – 2

$$\widehat{\mathbf{P}}(\boldsymbol{x}) = \left(G_1^1[x_1] \otimes \ldots \otimes G_1^m[x_1]\right) \ldots \left(G_n^1[x_n] \otimes \ldots \otimes G_n^m[x_n]\right).$$

Обозначим: $A_i[x_i] = G_i^1[x_i] \otimes \ldots \otimes G_i^m[x_i]$.

Итак,

$$Z = \sum_{\boldsymbol{x}} \widehat{\mathbf{P}}(\boldsymbol{x}) = \sum_{x_{\mathbf{1}}, \ldots, x_{\boldsymbol{n}}} A_1[x_1] \ldots A_n[x_n] =$$

$$= \left(\sum_{x_{\mathbf{1}}} A_1[x_1]\right) \ldots \left(\sum_{x_{\boldsymbol{n}}} A_n[x_n]\right) = B_1 \ldots B_n,$$

где

$$B_i = \sum_{x_{\boldsymbol{i}}=1}^{d_{\boldsymbol{i}}} A_i[x_i].$$

Алгоритм оценки нормировочной константы – 3

$$Z = B_1 \dots B_n,$$

Алгоритм:

1: Представить каждый отдельный фактор $\boldsymbol{\Psi}_\ell(\boldsymbol{x}^\ell)$ в TT-формате
2: Инициализировать $\boldsymbol{f}_{n+1} := 1$
3: **for** $i := n$ **downto** 1 **do**
4: Найти $A_i[x_i] := \otimes_{\ell=1}^{m} G_i^\ell[x_i]$
5: $B_i := \sum_{x_i=1}^{d_i} A_i[x_i]$
6: $\boldsymbol{f}_i := \text{round}(B_i \boldsymbol{f}_{i+1}, \varepsilon)$
7: **end for**
8: $\widehat{Z} := \boldsymbol{f}_1$

Теоретические гарантии

Теорема. Ошибка алгоритма оценки нормировочной константы ограничена следующим образом:

$$\left|Z - \widehat{Z}\right| \leq \|B_1\|_2 \dots \|B_n\|_2 \left((1+\varepsilon)^{n-1} - 1\right),$$

где ε — точность округления.

Маргинальные распределения

Наш подход позволяет получать маргинальные распределения:

$$\hat{p}_i(x_i) = \sum_{\boldsymbol{x}\setminus x_i} \widehat{\mathbf{P}}(\boldsymbol{x}) = \sum_{\boldsymbol{x}\setminus x_i} A_1[x_1] \dots A_n[x_n] =$$

$$= B_1 \dots B_{i-1}\, A_i[x_i]\, B_{i+1} \dots B_n,$$

где

$$B_i = \sum_{x_i=1}^{d_i} A_i[x_i].$$

Поиск минимума энергии

Сравнение с алгоритмом TRW-S на задачах из OpenGM:

Задача	n	d	Результат
geo-surf-3/gm6	320	3	48.41%
geo-surf-3/gm20	348	3	95.83%
geo-surf-3/gm203	187	3	98.69%
geo-surf-7/gm11	125	7	1 769.65%
matching/matching1	19	19	135.47%

Нормировочная константа

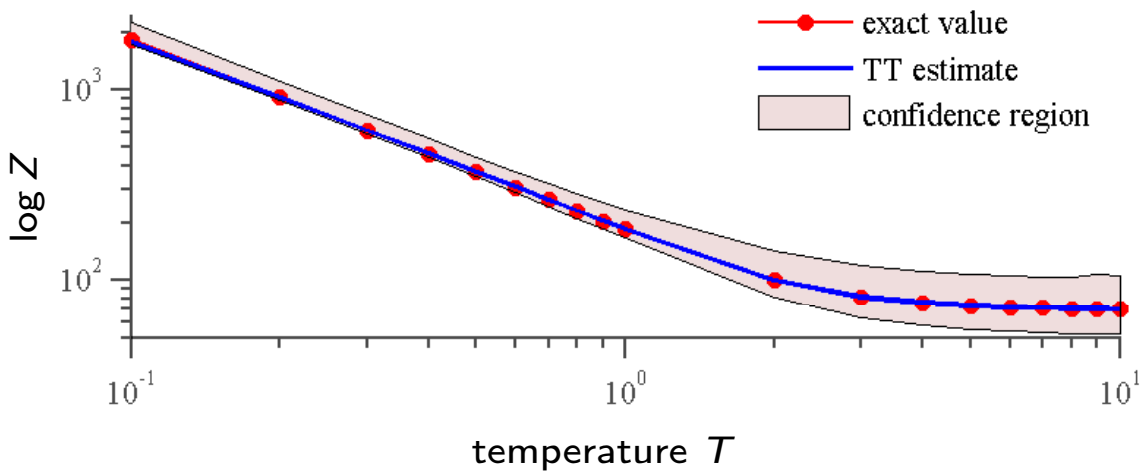
Модель Изинга:

$$\mathbf{E}(\boldsymbol{x}) = -\frac{1}{T}\left(\sum_{i=1}^{n} x_i h_i + \sum_{i,j} c_{ij} x_i x_j\right),$$

где $x_i \in \{-1, 1\}$.

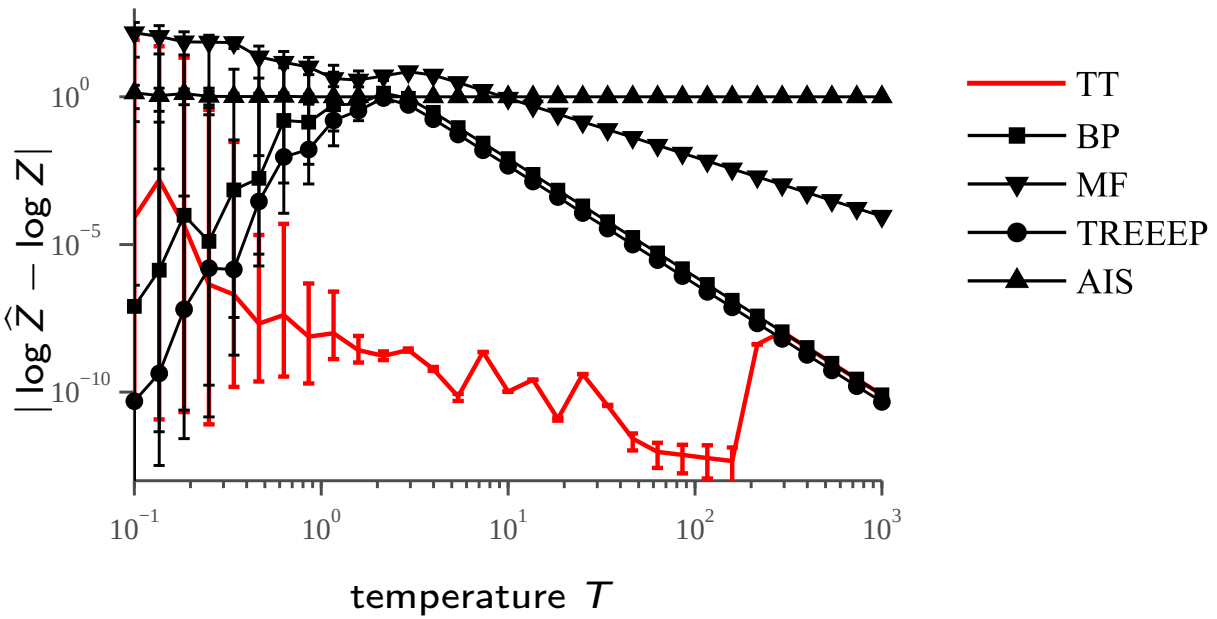
- Унарные коэффициенты $h_i \sim U[-1, 1]$.
- Во всех экспериментах размер решетки 10×10.

Точность оценки нормировочной константы



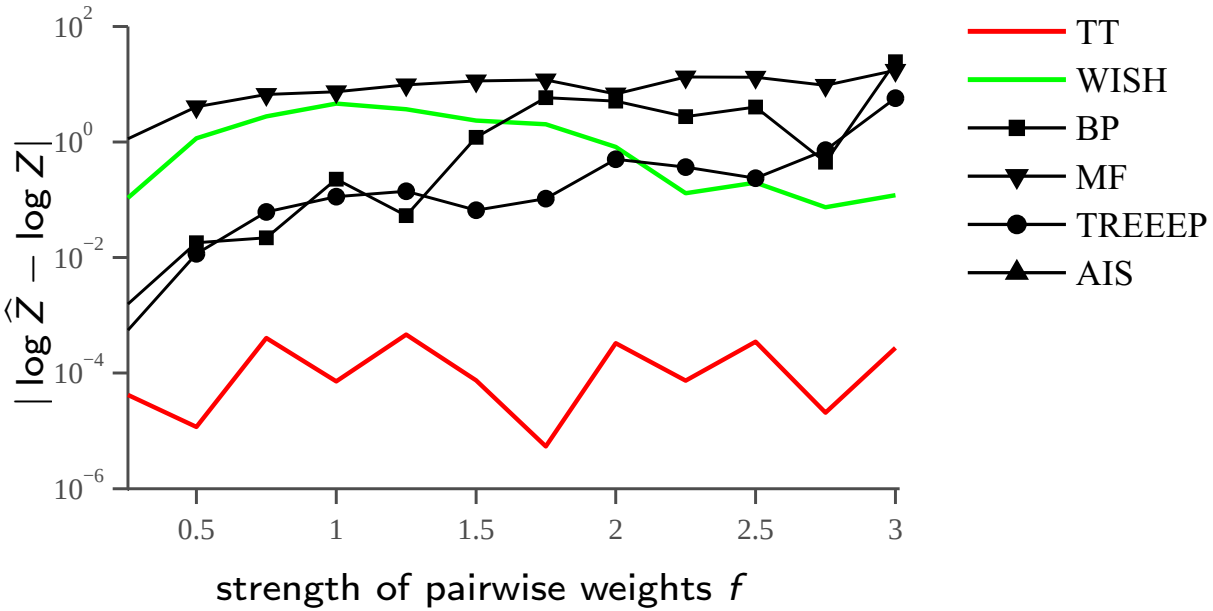
Теоретические гарантии на точность оценки нормировочной константы для модели Изинга ($c_{ij} = 1$), усреднение по 10 моделям.

Сравнение с аналогами



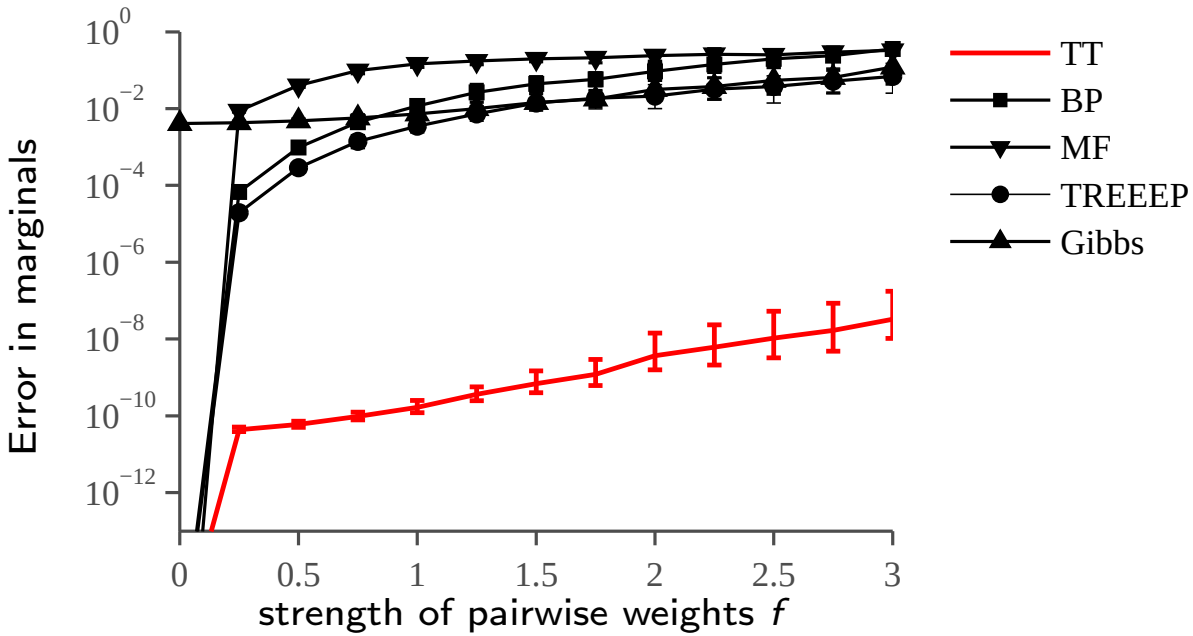
Сравнение на модели Изинга ($c_{ij} = 1$) с алгоритмами из библиотеки LibDAI [Mooij, 2010], усреднение по 50 моделям.

Сравнение с WISH



Сравнение на модели Изинга ($T = 1$, $c_{ij} \sim U[-f, f]$) с алгоритмом WISH [Ermon et al., 2013].

Маргинальные распределения



Сравнение на моделях Изинга ($T = 1$, $c_{ij} \sim U[-f, f]$) с алгоритмами из библиотеки LibDAI, усреднение по 50 моделям. Приведены ошибки оценки вероятности класса «-1».

Заключение

Вклад:

- Предложен точный алгоритм представления тензора энергии $\mathbf{E}$ в TT-формате с верхними оценками на TT-ранги;
- Продемонстрирована применимость TT-Toolbox в задаче MAP;
- Предложен алгоритм оценки нормировочной константы и маргинальных распределений;
- Проведены эксперименты по сравнению методов с аналогами.

Будущие направления работы:

- Аналитические формулы TT-разложения для потенциалов высокого порядка;
- Масштабирование метода оценки нормировочной константы на б**о**льшие задачи;
- Применение метода для обучения MRF.

Ссылки I


Dolgov, S. V. and Savostyanov, D. V. (2013).
Alternating minimal energy methods for linear systems in higher dimensions. Part II: Faster algorithm and application to nonsymmetric systems.
arXiv preprint 1304.1222.


Ermon, S., Gomes, C., Sabharwal, A., and Selman, B. (2013).
Taming the curse of dimensionality: Discrete integration by hashing and optimization.
In *International Conference on Machine Learning (ICML)*.


Mooij, J. M. (2010).
libDAI: A free and open source C++ library for discrete approximate inference in graphical models.
Journal of Machine Learning Research, 11:2169–2173.


Neal, R. (2001).
Annealed importance sampling.
Statistics and Computing, 11:125–139.

Ссылки II


Oseledets, I. V. (2011).
Tensor-Train decomposition.
SIAM J. Scientific Computing, 33(5):2295–2317.


Savostyanov, D. V. and Oseledets, I. V. (2011).
Fast adaptive interpolation of multi-dimensional arrays in tensor train format.
In *Proceedings of 7th International Workshop on Multidimensional Systems (nDS)*.