Mathematical Forecasting Methods

Лекция 15

МФТИ

Весна, 2024

Напоминание: CP-разложение и разложение Такера

$\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ - тензор N-ого порядка.
Пусть $I = \max\{I_1, ..., I_N\}$, тогда число элементов $\underline{\mathbf{X}}$ есть $O(I^N)$.

- Каноническое разложение:

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{b}_r^{(1)} \circ \mathbf{b}_r^{(2)} \circ ... \circ \mathbf{b}_r^{(N)} = [\![\underline{\mathbf{I}}; \mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \dots, \mathbf{B}^{(N)}]\!]$$

  Число элементов разложения: $O(NIR)$, где минимальное R – канонический ранг тензора, причем $R \leq I^{N-1}$.

- Разложение Такера, $\underline{\mathbf{G}} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}$ – центральный тензор N-ого порядка:

$$\underline{\mathbf{X}} = [\![\underline{\mathbf{G}}; \mathbf{U}^{(1)}, \dots, \mathbf{U}^{(N)}]\!] := \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \cdots \times_N \mathbf{U}^{(N)}$$

  Число элементов разложения: $O(NIR + R^N)$, где $R = \max\{R_1, ..., R_N\}$.

Напоминание: свёртка тензоров

- (N, 1)-свёртка (contraction) тензоров $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ и $\underline{\mathbf{B}} \in \mathbb{R}^{J_1 \times \cdots \times J_M}$, где обязательно $I_N = J_1$, даёт тензор $\underline{\mathbf{C}} \in \mathbb{R}^{I_1 \times \cdots \times I_{N-1} \times J_2 \times \cdots \times J_M}$ с элементами:

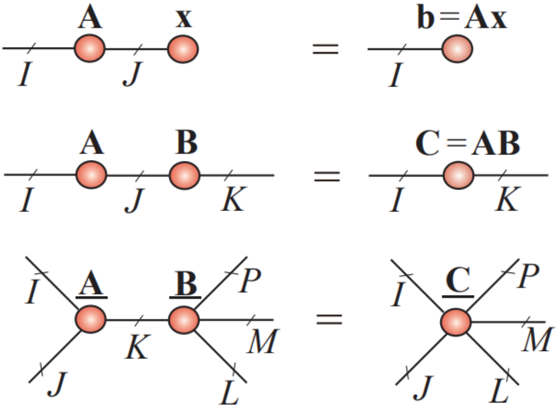
$$\underline{\mathbf{C}} = \underline{\mathbf{A}} \times_N^1 \underline{\mathbf{B}} = \underline{\mathbf{A}} \times^1 \underline{\mathbf{B}} = \underline{\mathbf{A}} \bullet \underline{\mathbf{B}}$$

$$c_{i_1,\ldots,i_{N-1},j_2,\ldots,j_M} = \sum_{i_N=1}^{I_N} a_{i_1,\ldots,i_{N-1},i_N} b_{i_N,j_2,\ldots,j_M}$$

Замечания:

- Свёртку можно осуществлять вдоль произвольных мод исходных тензоров: $\underline{\mathbf{C}} = \underline{\mathbf{A}} \times_n^m \underline{\mathbf{B}}$, требование совпадения соответствующих размерностей сохраняется.
- Свёртку также можно осуществлять по нескольким модам тензоров-участников одновременно: операция определяется двумя упорядоченными наборами индексов мод первого и второго тензоров, с условием совпадения соответствующих размерностей в каждой паре (см. лекцию 10).

Напоминание: графическое представление тензоров



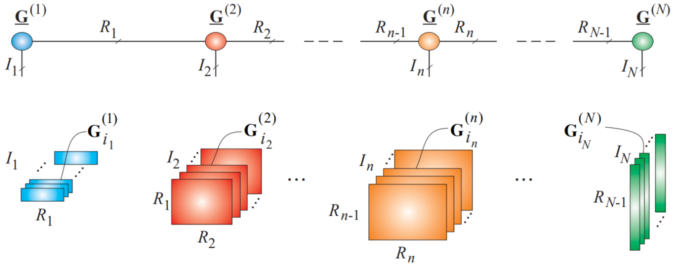
$$\sum_{k=1}^{K} a_{i,j,k}\, b_{k,l,m,p} = c_{i,j,l,m,p}$$

Разложение Tensor Train

- Тензор $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ представляется в виде последовательных свёрток тензоров:

$$\underline{\mathbf{X}} = \underline{\mathbf{G}}^{(1)} \times^1 \underline{\mathbf{G}}^{(2)} \times^1 \dots \times^1 \underline{\mathbf{G}}^{(N)} = \underline{\mathbf{G}}^{(1)} \bullet \underline{\mathbf{G}}^{(2)} \bullet \dots \bullet \underline{\mathbf{G}}^{(N)}$$

- $\underline{\mathbf{G}}^{(n)} \in \mathbb{R}^{R_{n-1} \times I_n \times R_n}, \;\; n \in \overline{1, N}$ – тензоры 3-его порядка.
- Полагаем $R_0 = R_N = 1$.
- При этом говорят, что тензор $\underline{\mathbf{X}}$ представлен в виде TT-разложения с TT-рангом $\mathbf{r}_{TT} = \{R_1, \dots, R_{N-1}\}$.



Преимущества Tensor Train

- Число элементов разложения: $O(NIR^2)$, где $R = \max\{R_1, ..., R_{N-1}\}$.
- Удобство применения операций для тензоров в ТТ-формате, например, сложение:

$$\underline{\mathbf{X}} = \underline{\mathbf{X}}^{(1)} \times^1 \underline{\mathbf{X}}^{(2)} \times^1 \cdots \times^1 \underline{\mathbf{X}}^{(N)} \quad \text{(ТТ-разложение)}$$

$$\underline{\mathbf{Y}} = \underline{\mathbf{Y}}^{(1)} \times^1 \underline{\mathbf{Y}}^{(2)} \times^1 \cdots \times^1 \underline{\mathbf{Y}}^{(N)} \quad \text{(ТТ-разложение)}$$

$$\underline{\mathbf{Z}} = \underline{\mathbf{X}} + \underline{\mathbf{Y}} = \underline{\mathbf{Z}}^{(1)} \times^1 \underline{\mathbf{Z}}^{(2)} \times^1 \cdots \times^1 \underline{\mathbf{Z}}^{(N)}$$

$$\mathbf{r}_{TT}(\underline{\mathbf{Z}}) = \mathbf{r}_{TT}(\underline{\mathbf{X}}) + \mathbf{r}_{TT}(\underline{\mathbf{Y}}), \quad \underline{\mathbf{Z}}^{(n)} \in \mathbb{R}^{(R^X_{n-1} + R^Y_{n-1}) \times I_n \times (R^X_n + R^Y_n)}$$

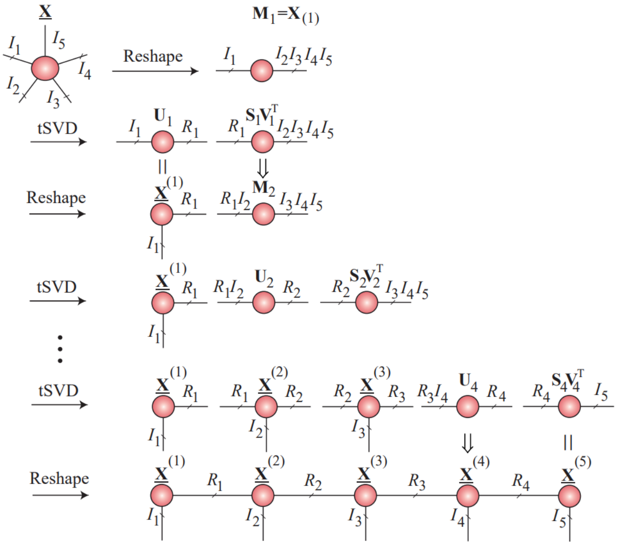
$$\underline{\mathbf{Z}}^{(n)}_{:,i_n,:} = \begin{bmatrix} \underline{\mathbf{X}}^{(n)}_{:,i_n,:} & \mathbf{0} \\ \mathbf{0} & \underline{\mathbf{Y}}^{(n)}_{:,i_n,:} \end{bmatrix}, \; n \in \overline{2, N-1}$$

$$\underline{\mathbf{Z}}^{(1)}_{:,i_n,:} = \begin{bmatrix} \underline{\mathbf{X}}^{(1)}_{:,i_n,:} & \underline{\mathbf{Y}}^{(1)}_{:,i_n,:} \end{bmatrix}, \quad \underline{\mathbf{Z}}^{(N)}_{:,i_n,:} = \begin{bmatrix} \underline{\mathbf{X}}^{(N)}_{:,i_n,:} \\ \underline{\mathbf{Y}}^{(N)}_{:,i_n,:} \end{bmatrix}$$

Другие операции в TT-формате

Operation	TT-cores
$\underline{\mathbf{Z}} = \underline{\mathbf{X}} + \underline{\mathbf{Y}} = \left(\underline{\mathbf{X}}^{(1)} \oplus_2 \underline{\mathbf{Y}}^{(1)}\right) \times^1 \left(\underline{\mathbf{X}}^{(2)} \oplus_2 \underline{\mathbf{Y}}^{(2)}\right) \times^1 \cdots \times^1 \left(\underline{\mathbf{X}}^{(N)} \oplus_2 \underline{\mathbf{Y}}^{(N)}\right)$ $\underline{\mathbf{Z}}^{(n)} = \underline{\mathbf{X}}^{(n)} \oplus_2 \underline{\mathbf{Y}}^{(n)}$, with TT core slices $\mathbf{Z}^{(n)}_{i_n} = \mathbf{X}^{(n)}_{i_n} \oplus \mathbf{Y}^{(n)}_{i_n}$, $(I_n = J_n = K_n,\ \forall n)$	
$\underline{\mathbf{Z}} = \underline{\mathbf{X}} \oplus \underline{\mathbf{Y}} = \left(\underline{\mathbf{X}}^{(1)} \oplus \underline{\mathbf{Y}}^{(1)}\right) \times^1 \left(\underline{\mathbf{X}}^{(2)} \oplus \underline{\mathbf{Y}}^{(2)}\right) \times^1 \cdots \times^1 \left(\underline{\mathbf{X}}^{(N)} \oplus \underline{\mathbf{Y}}^{(N)}\right)$	
$\underline{\mathbf{Z}} = \underline{\mathbf{X}} \circledast \underline{\mathbf{Y}} = \left(\underline{\mathbf{X}}^{(1)} \odot_2 \underline{\mathbf{Y}}^{(1)}\right) \times^1 \left(\underline{\mathbf{X}}^{(2)} \odot_2 \underline{\mathbf{Y}}^{(2)}\right) \times^1 \cdots \times^1 \left(\underline{\mathbf{X}}^{(N)} \odot_2 \underline{\mathbf{Y}}^{(N)}\right)$ $\underline{\mathbf{Z}}^{(n)} = \underline{\mathbf{X}}^{(n)} \odot_2 \underline{\mathbf{Y}}^{(n)}$, with TT core slices $\mathbf{Z}^{(n)}_{i_n} = \mathbf{X}^{(n)}_{i_n} \otimes \mathbf{Y}^{(n)}_{i_n}$, $(I_n = J_n = K_n,\ \forall n)$	
$\underline{\mathbf{Z}} = \underline{\mathbf{X}} \otimes \underline{\mathbf{Y}} = \left(\underline{\mathbf{X}}^{(1)} \otimes \underline{\mathbf{Y}}^{(1)}\right) \times^1 \left(\underline{\mathbf{X}}^{(2)} \otimes \underline{\mathbf{Y}}^{(2)}\right) \times^1 \cdots \times^1 \left(\underline{\mathbf{X}}^{(N)} \otimes \underline{\mathbf{Y}}^{(N)}\right)$ $\underline{\mathbf{Z}}^{(n)} = \underline{\mathbf{X}}^{(n)} \otimes \underline{\mathbf{Y}}^{(n)}$, with TT core slices $\mathbf{Z}^{(n)}_{k_n} = \mathbf{X}^{(n)}_{i_n} \otimes \mathbf{Y}^{(n)}_{j_n}$ $\quad (k_n = \overline{i_n j_n})$	
$\underline{\mathbf{Z}} = \underline{\mathbf{X}} * \underline{\mathbf{Y}} = (\underline{\mathbf{X}}^{(1)} \boxdot_2 \underline{\mathbf{Y}}^{(1)}) \times^1 \cdots \times^1 (\underline{\mathbf{X}}^{(N)} \boxdot_2 \underline{\mathbf{Y}}^{(N)})$ $\underline{\mathbf{Z}}^{(n)} = \underline{\mathbf{X}}^{(n)} \boxdot_2 \underline{\mathbf{Y}}^{(n)} \in \mathbb{R}^{(R_{n-1}Q_{n-1}) \times (I_n + J_n - 1) \times (R_n Q_n)}$, with vectors $\underline{\mathbf{Z}}^{(n)}(s_{n-1}, :, s_n) = \underline{\mathbf{X}}^{(n)}(r_{n-1}, :, r_n) * \underline{\mathbf{Y}}^{(n)}(q_{n-1}, :, q_n) \in \mathbb{R}^{(I_n + J_n - 1)}$ for $s_n = 1, 2, \ldots, R_n Q_n$ and $n = 1, 2, \ldots, N$, $R_0 = R_N = 1$.	
$\underline{\mathbf{Z}} = \underline{\mathbf{X}} \times_n \mathbf{A} = \underline{\mathbf{X}}^{(1)} \times^1 \cdots \times^1 \underline{\mathbf{X}}^{(n-1)} \times^1 \left(\underline{\mathbf{X}}^{(n)} \times_2 \mathbf{A}\right) \times^1 \underline{\mathbf{X}}^{(n+1)} \times^1 \cdots \times^1 \underline{\mathbf{X}}^{(N)}$	

Алгоритм TT-SVD Decomposition



Алгоритм TT-SVD Decomposition

Input: тензор $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ и ошибка аппроксимации ε.
Output: Приближенное представление тензора в формате TT $\{\underline{\mathbf{G}}^{(n)}\}$, такое что $\|\underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}}\|_F \le \varepsilon$.

1. $\mathbf{M}_1 = \underline{\mathbf{X}}_{(1)}$ — развертка по 1-ой моде и $R_0 = R_N = 1$.
2. **for** $n = 1$ **to** $N-1$ **do**
3. $\quad \mathbf{U}^{(n)}, \mathbf{S}^{(n)}, \mathbf{V}^{(n)} = \text{truncated\_SVD}\left(\mathbf{M}_n, \frac{\varepsilon}{\sqrt{N-1}}\right)$.
4. $\quad R_n = \text{rank}(\mathbf{U}^{(n)})$.
5. $\quad \underline{\mathbf{G}}^{(n)} = \text{reshape}(\mathbf{U}^{(n)}, [R_{n-1}, I_n, R_n])$ — преобразование в фактор-тензор.
6. $\quad \mathbf{M}_{n+1} = \text{reshape}(\mathbf{S}^{(n)} \mathbf{V}^{(n)\mathsf{T}}, [R_n I_{n+1}, \prod_{p=n+2}^{N} I_p])$ — преобразование в матрицу для следующей итерации.
7. **end for**
8. $\underline{\mathbf{G}}^{(N)} = \text{reshape}(\mathbf{U}^{(N)}, [R_{N-1}, I_N, R_N])$ — преобразование в последний фактор-тензор.
9. **return** $\hat{\underline{\mathbf{X}}} = \underline{\mathbf{G}}^{(1)} \times^1 \cdots \times^1 \underline{\mathbf{G}}^{(N)}$

Резюме

- Ранг канонического разложения может быть очень велик, а число элементов в разложении Такера всё ещё имеет показательную зависимость от порядка исходного тензора.
- Решить эту проблему можно, позволив факторам разложения быть тензорами порядка больше двух.
- Одно из классических разложений такого типа – разложение Tensor Train с фактор-тензорами 3-его порядка.
- В Tensor-Train формате удобно осуществлять многие математические операции над тензорами.
- Алгоритм TT-SVD позволяет приближенно найти TT-разложение исходного тензора.