Online Nonnegative Matrix Factorization with Temporal Affinity

Miao Cheng

Email: *mcheng@mailbox.gxnu.edu.cn*
School of Computer Science
Guangxi Normal University
ICSIP'22, Suzhou, China

Background

- The well-known Nonnegative Matrix Factorization (NMF) method

$$\begin{aligned} X \quad & \approx AU \\ s.t. \quad & A \geq 0, U \geq 0 \end{aligned} \tag{1}$$

- Online Nonnegative Matrix Factorization (ONMF) attempts to decompose the nonnegative data stream y_i, $i = 1, 2, \cdots, m$ with sequential learning.

Background

- The exisiting ONMF methods are hardly to reserve the nonnegative patterns during sequential learning. $(\surd)$
- The calculation efficiency is necessary to be held during temporal learning of ONMF. $(\surd)$
- As a result, there posts more sequential demands on nonnegative learning.

Related Work

The most representative methods can be listed as below.

- INMF [S. S. Bucak, et. al., 2007]
- Wang's ONMF [F. Wang, et. al., 2011]
- Itakura-Saito divergence based ONMF [A. Lef´evre, et. al., 2011]
- Zhao's ONMF [R. Zhao et. al., 2016]

Related Work

- The INMF [S. S. Bucak, et. al., 2007] explains ONMF as an incremental subspace learning problem
- Optimizes the fusion of temporal objectives with penalty during sequential learning, such as

$$F_{m+1} = (1-\alpha)\, F_m + \alpha \sum_{i=1}^{n} \left(v_i - \sum_{a=1}^{r} W_{ia} h_a \right)^2 . \tag{2}$$

Related Work

- The Wang's ONMF [F. Wang, et. al., 2011] attempts to optimize sequential learning by preserving the characteristics of gradient descent of factorizations during temporal learning.
- Two optimization methods:
  - **One Pass ONMF:** The factorizations need to be recomputed at each time slice.
  - **MultiPass ONMF:** The factorizations can be updated using the results obtained in the previous pass.

Related Work

- The Itakura-Saito divergence based ONMF [A. Lef´evre, et. al., 2011] is an online learning extension of Itakura-Saito divergence based NMF.
- The optimization of Itakura-Saito NMF is a simplified advancement of KL divergence based NMF.

Related Work

- The Zhao's ONMF [R. Zhao et. al., 2016] is an online learning extension with calculation of residuals of nonnegative optimization.
- Particularly, the efficiency can be saved with the summarized calculation equations, which is derived from Wang's idea.

Temporal Learning

- Temporal learning seeks for the ideal explanation of data stream with respect to different time slices,

$$x_i \mapsto y_i, \;\; i = 1, 2, \cdots, t. \tag{3}$$

- Normally, the optimized learning among sequential time slices can enhance learning ability, and accumulated objective can be defined as

$$J_t\left(x_t\right) \leftarrow \sum_{i=1}^{t-1} a_i \cdot J_t\left(x_i\right) + a_t \cdot J_t\left(x_t\right). \tag{4}$$

Main Idea

- As a **common sense**, the online learning of NMF can be defined as a reconstruction of sequential objectives,

$$J_{t-1\to t} = \arg \min_{\substack{B_{t-1}, v_{t-1} \\ B_t, v_t}} \|y_{t-1} - B_{t-1} v_{t-1}\|^2 + \|y_t - B_t v_t\|^2 \tag{5}$$

Why not try the forward one directly ?

- With **triangle inequations**, the objective function can be further referred as a objective of temporal learning,

$$J_{t-1\to t} = \arg\min_{\substack{B_{t-1}, v_{t-1} \\ B_t, v_t}} (y_{t-1} - B_{t-1} v_{t-1})^T (y_t - B_t v_t) \tag{6}$$

- **The proposed method attempts to optimize temporal learning of NMF corresponding to reconstruction.**

Why not try the forward one directly ?

- With **triangle inequations**, the objective function can be further simplified as

$$J_{t-1\to t} = \arg\min_{\substack{B_{t-1},v_{t-1}\\ B_t,v_t}} \left(y_{t-1} - B_{t-1}v_{t-1}\right)^T \left(y_t - B_t v_t\right) \tag{4}$$

(×) It is not enough !

Why not try the forward one directly ?

Two issues:

- Learning stability
- Disadjusted learning

Why not try the forward one directly ?

Two issues:

- Learning stability
- Disadjusted learning

($\surd$) **The smooth affinity is brought into temporal learning**

Why not try the forward one directly ?

- The smooth of affinity is brought into temporal learning:

$$J_{t-1\to t} = \arg \min_{\substack{B_{t-1}, v_{t-1} \\ B_t, v_t}} \left(y_{t-1} - B_{t-1}v_{t-1}\right)^T M \left(y_t - B_t v_t\right) \tag{7}$$

- The appended M can **smooth** the nonnegative learning of sequential data, and **stable** performance is promised to be achieved.

Temporal Nonnegative Learning

- Associated with sequential learning of successive objectives in stream, it is to solve

$$
\begin{aligned}
J_{t-1\to t} = \; & \arg \min_{\substack{B_{t-1}, v_{t-1} \\ B_t, v_t}} (y_t - B_t v_t)^T M (y_t - B_t v_t) + \\
& (y_{t-1} - B_{t-1} v_{t-1})^T M (y_t - B_t v_t) + \\
& (y_{t-1} - B_{t-1} v_{t-1})^T M (y_{t-1} - B_{t-1} v_{t-1})
\end{aligned} \tag{8}
$$

Temporal Nonnegative Learning

- Accordingly, the derivatives of each items with respect to Eq. (8) can be calculated as

$$\begin{aligned}\frac{\partial J_t}{\partial B_{t-1}} = & -2M^T y_{t-1} v_{t-1}^T + 2MB_{t-1} v_{t-1} v_{t-1}^T \\ & -M y_t v_{t-1}^T + MB_t v_t v_{t-1}^T\end{aligned} \tag{9}$$

$$\begin{aligned}\frac{\partial J_t}{\partial B_t} = & -2M^T y_t v_t^T + 2MB_t v_t v_t^T \\ & -M^T y_{t-1} v_t^T + M^T B_{t-1} v_{t-1} v_t^T\end{aligned} \tag{10}$$

$$\begin{aligned}\frac{\partial J_t}{\partial v_t} = & -2B_t^T M y_t + 2B_t^T MB_t v_t \\ & -B_t^T M^T y_{t-1} + B_t^T M^T B_{t-1} v_{t-1}.\end{aligned} \tag{11}$$

Temporal Nonnegative Learning

- The update rules with the complementary slackness condition of Karush-Kuhn-Tucker (KKT) conditions,

$$B_{t-1} \leftarrow B_{t-1} \cdot \frac{2M^T y_{t-1} v_{t-1}^T + M y_t v_{t-1}^T}{2MB_{t-1} v_{t-1} v_{t-1}^T + MB_t v_t v_{t-1}^T} \tag{12}$$

$$B_t \leftarrow B_t \cdot \frac{2M^T y_t v_t^T + M^T y_{t-1} v_t^T}{2MB_t v_t v_t^T + M^T B_{t-1} v_{t-1} v_t^T} \tag{13}$$

$$v_t \leftarrow v_t \cdot \frac{2B_t^T M y_t + B_t^T M^T y_{t-1}}{2B_t^T M B_t v_t + B_t^T M^T B_{t-1} v_{t-1}}. \tag{14}$$

- **Tips: B_{t-1} will not be updated with the newly calculated one, as it actually works as a corresponding variant between time slices.**

Temporal Nonnegative Learning

- The functions with temporally increasing tendency are competent to optimized fusion of B_t, for instance,

$$B_t = \frac{1}{\sum_{i=1}^{t} a^{\frac{i}{t}}} \sum_{i=1}^{t} a^{\frac{i}{t}} B_i \tag{15}$$

$$B_t = \frac{1}{\sum_{i=1}^{t} Sin\left(\frac{i\pi}{2t}\right)} \sum_{i=1}^{t} Sin\left(\frac{i\pi}{2t}\right) B_i \tag{16}$$

Experiment

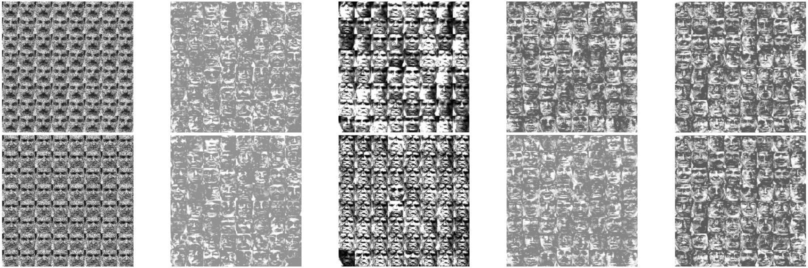
Several state-of-the-art methods

- Baseline: independent online learning of each time slice
- INMF [Bucak '07]
- ONMF [Wang '11]
- ONMFO [Zhao '16]
- The proposed TANMF

Experimental Parameters

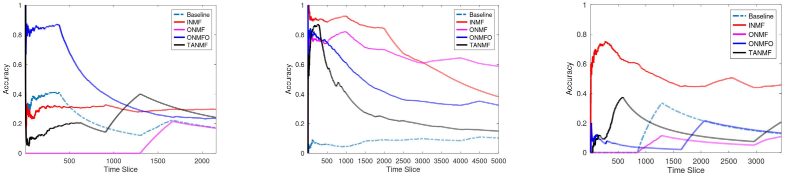
- For the ONMF method [Wang11ONMF], the diagonal approximation is adopted to online learning of sequential data.
- For each algorithm, the maximum iteration of nonnegative learning is set to be 100 times, memorized update of TANMF is set to be 20 slices.

First Experiment



The obtained first 64 factorizations of different algorithms on the Yale database in the 10th (above) and 90th (bottom) time slices. (a) Baseline (b) INMF (c) ONMF (d) ONMFO (e) TANMF.

Second Experiment



The online learning results of different algorithms on three data sets. (a) Flower (b) MNIST (c) Natural Images.

Conclusion

- In this work, a novel approach to online NMF is proposed to perform sequential nonnegative learning, while temporal affinity of factorizations are adopted to enhance the learning abilities.
- More stable and robust performance can be achieved by the proposed method.
- Experiments on several data sets demonstrate that, the proposed method is able to give fascinating results compared with the state-of-the-art methods.

Thank You

@ ICSIP'22, Suzhou, China