Extracting Highly Effective Features for Supervised Learning via Simultaneous Tensor Factorization

Sunny Verma,† Wei Liu,‡ Chen Wang,† Liming Zhu‡

†Advanced Analytics Institute, University of Technology Sydney, Sydney, Australia
‡Data61, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Sydney, Australia
Sunny.Verma@student.uts.edu.au, Wei.Liu@uts.edu.au, Chen.Wang@data61.csiro.au, Liming.Zhu@data61.csiro.au

Abstract

Real world data is usually generated over multiple time periods associated with multiple labels, which can be represented as multiple labeled tensor sequences. These sequences are linked together, sharing some common features while exhibiting their own unique features. Conventional tensor factorization techniques are limited to extract either common or unique features, but not both simultaneously. However, both types of these features are important in many machine learning systems as they inherently affect the systems’ performance. In this paper, we propose a novel supervised tensor factorization technique which simultaneously extracts common and unique features. Classification results using features extracted by our method on CIFAR-10 database achieves significantly better performance over other factorization methods, illustrating the effectiveness of the proposed technique.

Related Work and Tensor Notations

Tensors are higher order generalizations of matrices denoted in this paper by boldface Euler script letters $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$. An $N$ mode (dimensions in matrix) tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ can be rearranged as a matrix in any chosen mode $n$, denoted in boldface capital letter $X_{(n)}$. For a detailed review on tensors and related factorization literature, refer to Kolda and Bader.

Acar, Kolda, and Dunlavy were the first to propose extraction of common features shared among multiple data sources. In their work, the authors proposed joint factorization of a tensor with a matrix sharing common features on a single identical mode - coupled matrix and tensor factorization (CMTF). However, in their work, the authors did not address these two issues: 1) how to extract common features from more than one identical mode, and more importantly 2) how to extract unique features from the same shared identical mode. These challenging issues are addressed by our method introduced in the next section, which demonstrates the power of having unique discriminative features (Ristaniški, Liu, and Bailey 2013).

Proposed CUTF

CUTF is developed using Higher-Order Orthogonal Iteration (HOOI) technique (Liu et al. 2015). HOOI is a generalization of matrix SVD technique and is developed for factorizing single tensors. It decomposes tensor $\mathcal{X} \approx \langle \mathcal{G}, \mathbf{A}^{(1)} \times_1 \mathbf{A}^{(2)} \times_2 \ldots \times_N \mathbf{A}^{(N)} \rangle$, mathematically $\mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \ldots \times_N \mathbf{A}^{(N)}$ representing sequential multiplication of a tensor with a matrix in $i^{th}$-mode ($1 \leq i \leq N$). Here, $\mathcal{G}$ can be thought as compressed version of $\mathcal{X}$ and $\mathbf{A}^{(i)}$ represents low-rank factor matrices of $i^{th}$-mode in tensor $\mathcal{X}$.

Without loss of generality, we focus on two 3-mode tensors $\mathcal{X}$ and $\mathcal{Y} \in \text{Class} \{+1, -1\}$, sharing common features in their first mode. Denote $\mathbf{W}$ as the common features shared among tensors in their first mode and, denote...
V and S as the remaining unique features of tensors in the same mode, and denote U^j and K^i as the factors of other modes of X and Y. Simply, (W|V) and (W|S) represents factor matrices of the tensors X and Y in their first mode respectively. Our objective is to jointly factorize X and Y to obtain their low rank approximations, simultaneously extracting their common (W) and unique (V, S) features:

\[ \text{obj} = \min \left[ \|X - [S X; (W|V), U^{(2)}, U^{(3)}]\|_F + \|Y - [S Y; (W|S), K^{(2)}, K^{(3)}]\|_F \right]. \]

The complete procedure of solving our objective function is presented in Algorithm 1.

**Algorithm 1: Common and Unique Tensor Factorization**

1. Input (X, Y, R_x, R_y, R_s, Master)
2. Output [S X; (W|V), U^{(2)}, U^{(3)}, S Y; (W|S), K^{(2)}, K^{(3)}]
3. \( \text{U}^{(1)} \leftarrow \text{R}_x \), left singular vectors of \( \text{X}_{(1)} \) \( i = 2, 3 \)
4. \( \text{K}^{(1)} \leftarrow \text{R}_s \), left singular vectors of \( \text{Y}_{(1)} \) \( i = 2, 3 \)
5. \( \text{V} \leftarrow \left[ \text{R}_x / 2 \right] \) left singular vectors of \( \text{X}_{(1)} \)
6. \( \text{S} \leftarrow \left[ \text{R}_x / 2 \right] \) left singular vectors of \( \text{Y}_{(1)} \)
7. \( \text{G} \leftarrow \text{X}; (\text{W}|\text{V})^T, (\text{U}^{(2)})^T, (\text{U}^{(3)})^T \)
8. \( \text{N} \leftarrow \text{Y}; (\text{W}|\text{S})^T, (\text{K}^{(2)})^T, (\text{K}^{(3)})^T \)
9. \( \text{K}^{(1)} \leftarrow \text{R}_x \), left singular vectors of \( \text{N}_{(1)} \)
10. while \( \text{obj} \) converges or Master exhausted do
11. for \( i = 2, 3 \) do
12. \( \text{M} \leftarrow \text{G}; (\text{W}|\text{V})^T, (\text{U}^{(1)})^T \)
13. \( \text{U}^{(1)} \leftarrow \text{R}_x \), left singular vectors of \( \text{M}_{(1)} \)
14. \( \text{N} \leftarrow \text{N}; (\text{W}|\text{S})^T, (\text{K}^{(1)})^T \)
15. \( \text{K}^{(1)} \leftarrow \text{R}_x \), left singular vectors of \( \text{N}_{(1)} \)
16. where \( j \in [2, 3], k \neq i \)
17. end for
18. \( \text{M} \leftarrow \text{G}; (\text{U}^{(2)})^T, (\text{U}^{(3)})^T \)
19. \( \text{N} \leftarrow \text{N}; (\text{K}^{(2)})^T, (\text{K}^{(3)})^T \)
20. \( \text{W} \leftarrow \left[ \text{R}_x / 2 \right] \) left singular vectors of \( \text{M}_{(1)} \)
21. \( \text{V} \leftarrow \left[ \text{R}_x / 2 \right] \) left singular vectors of \( \text{M}_{(1)} \)
22. \( \text{S} \leftarrow \left[ \text{R}_x / 2 \right] \) left singular vectors of \( \text{N}_{(1)} \)
23. \( \text{G} \leftarrow \text{X}; (\text{W}|\text{V})^T, (\text{U}^{(2)})^T, (\text{U}^{(3)})^T \)
24. \( \text{N} \leftarrow \text{Y}; (\text{W}|\text{S})^T, (\text{K}^{(2)})^T, (\text{K}^{(3)})^T \)
25. end while

**Experiments and Analysis**

To evaluate the significance of features extracted using CUTF, we utilized CIFAR-10 dataset (Krizhevsky and Hinton 2009), which consists of 60K RGB images of 32 × 32 pixels equally divided among 10 categories. For each label, we build a tensor of 4 modes: \( \text{RowPixels} \times \text{ColumnPixels} \times \text{color} \times \text{position} \). We randomly chose multiple pairs of binary categories from the database and extracted three different feature sets: 1) common (Com), 2) unique (Unq) and, 3) both common and unique (e.g., CUTF). Note that the Com is the same as the CMTF.

CUTF is implemented using Matlab tensor toolbox (Bader, Kolda, and others 2015). Extracted features are classified using linear Logistic Regression (LR) and SVM with polynomial kernel (SVM-Poly). To validate the superiority of CUTF, Friedman tests were performed on the classification results, and \( p \)-values are reported in the bottom of Table 1. These low \( p \)-values illustrate the statistical significance of our technique. Moreover, Fig-1 compares the accuracies obtained through SVM-Poly on different factorization ranks, demonstrating the advantages of the proposed method.

**Conclusion and Future Work**

In this research, we have proposed a novel supervised tensor factorization technique, which simultaneously extracts common and unique features. These features are ordered by their singular value significance with respect to multiple labeled tensor sequences. Experiments reported in this paper demonstrate huge potential of simultaneously extracting common and unique features. Our future work includes extending the proposed CUTF for sparse tensor factorizations.

**References**


