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 ordered *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.

Introduction and Motivation

In the real world, data is often acquired as a sequence of matrices rather than a single matrix. These matrices can be represented as multiple labeled tensor (multidimensional arrays) sequences (Lahat, Adali, and Jutten 2015). Due to the underlying data generation mechanism, these sequences are naturally linked together and share some *common* features. While at the same time, they also exhibit their own *unique* features. When one is faced with the scenario of extracting features from these multiple labeled tensors sequences, two common approaches are followed: a) concatenate all tensor instances and factorize them together, or b) factorize each tensor instance individually. However, both these approaches suffer information loss. The former approach is limited to extract *common* features, suffering the loss of unique features. While, the latter approach is limited to extract *unique* features, suffering the loss of common features. This is because conventional tensor factorization techniques are unsupervised i.e., tensor instances are factorized without considering its label (category). Hence, one is only able to extract either *common* or *unique* features, but not both simultaneously from multiple labeled tensor sequences.

To reduce this information loss in conventional factorization techniques, we propose a novel supervised tensor factorization technique called **C**ommon and **U**nique **T**ensor **F**actorization (**CUTF**). The proposed technique simultaneously extracts *common* and *unique* features from multiple labeled tensor sequences. Furthermore, the extracted features are ordered by their singular value significance, enabling feature re-utilization in several machine learning tasks.

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 $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \dots \times I_N}$ can be rearranged as a matrix in any chosen mode *n*, denoted in boldface capital letter $\mathbf{X}_{(n)}$. 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 addressed 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 (Ristanoski, 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 $\mathcal{X} \approx [\![\mathcal{G}; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, ..., \mathbf{A}^{(N)}]\!]$, mathematically $\mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} ... \times_N \mathbf{A}^{(N)}$ 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 $\mathbf{A}^{(i)s}$ represents low-rank factor matrices of i^{th}-mode in tensor $\mathcal{X}$. *HOOI* guarantees best rank-$(R_1, R_2, ..., R_N)$ approximation of tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \dots \times I_N}$ (Liu et al. 2014).

Without loss of generality, we focus on two 3-mode tensors $\mathcal{X}$ and $\mathcal{Y}$ $\in$ 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

$\mathbf{V}$ and $\mathbf{S}$ as the remaining unique features of tensors in the same mode, and denote $\mathbf{U}^i$ and $\mathbf{K}^i$ as the factors of other modes of $\mathcal{X}$ and $\mathcal{Y}$. Simply, $(\mathbf{W}|\mathbf{V})$ and $(\mathbf{W}|\mathbf{S})$ represents factor matrices of the tensors $\mathcal{X}$ and $\mathcal{Y}$ in their first mode respectively. Our objective is to jointly factorize $\mathcal{X}$ and $\mathcal{Y}$ to obtain their low rank approximations, simultaneously extracting their common $(\mathbf{W})$ and unique $(\mathbf{V}, \mathbf{S})$ features: $obj = \min [\ \| \mathcal{X} - [\![\mathcal{G}_\mathcal{X}; (\mathbf{W}|\mathbf{V}), \mathbf{U}^{(2)}, \mathbf{U}^{(3)}]\!] \|_F + \| \mathcal{Y} - [\![\mathcal{G}_\mathcal{Y}; (\mathbf{W}|\mathbf{S}), \mathbf{K}^{(2)}, \mathbf{K}^{(3)}]\!] \|_F\ ]$. The complete procedure of solving our objective function is presented in Algorithm 1.

Algorithm 1 **C**ommon and **U**nique **T**ensor **F**actorization

In $(\mathcal{X}, \mathcal{Y}, \mathbf{R_1}, \mathbf{R_2}, \mathbf{R_3}, \mathit{MaxIter})$
Out $[\mathcal{G}_\mathcal{X}, (\mathbf{W}|\mathbf{V}), \mathbf{U}^{(2)}, \mathbf{U}^{(3)}, \mathcal{G}_\mathcal{Y}, (\mathbf{W}|\mathbf{S}), \mathbf{K}^{(2)}, \mathbf{K}^{(3)}]$
$\mathbf{U}^{(i)} \leftarrow \mathbf{R}_i$ left singular vectors of $\mathbf{X}_{(i)}$ $i = 2, 3$
$\mathbf{K}^{(i)} \leftarrow \mathbf{R}_i$ left singular vectors of $\mathbf{Y}_{(i)}$ $i = 2, 3$
$\mathbf{W} \leftarrow \lceil \mathbf{R}_1/2 \rceil$ left singular vectors of $[\mathbf{X}_{(1)} \mathbf{Y}_{(1)}]$
$\mathbf{V} \leftarrow \lfloor \mathbf{R}_1/2 \rfloor$ left singular vectors of $\mathbf{X}_{(1)}$
$\mathbf{S} \leftarrow \lfloor \mathbf{R}_1/2 \rfloor$ left singular vectors of $\mathbf{Y}_{(1)}$
$\mathcal{G}_\mathcal{X} \leftarrow [\![\mathcal{X}; (\mathbf{W}|\mathbf{V})^T, (\mathbf{U}^{(2)})^T, (\mathbf{U}^{(3)})^T]\!]$
$\mathcal{G}_\mathcal{Y} \leftarrow [\![\mathcal{Y}; (\mathbf{W}|\mathbf{S})^T, (\mathbf{K}^{(2)})^T, (\mathbf{K}^{(3)})^T]\!]$
while obj converges or *MaxIter* exhausted **do**
  for $i = 2, 3$ **do**
    $\mathcal{M} \leftarrow [\![\mathcal{G}_\mathcal{X}; (\mathbf{W}|\mathbf{V})^T, (\mathbf{U}^{(j)})^T]\!]$
    $\mathbf{U}^{(j)} \leftarrow \mathbf{R}_j$ left singular vectors of $\mathbf{M}_{(j)}$
    $\mathcal{N} \leftarrow [\![\mathcal{G}_\mathcal{Y}; (\mathbf{W}|\mathbf{S})^T, (\mathbf{K}^{(j)})^T]\!]$
    $\mathbf{K}^{(j)} \leftarrow \mathbf{R}_j$ left singular vectors of $\mathbf{N}_{(j)}$
    where $j \in [2, 3]$ & $j \neq i$
  end for
  $\mathcal{M} \leftarrow [\![\mathcal{G}_\mathcal{X}; (\mathbf{U}^{(2)})^T, (\mathbf{U}^{(3)})^T]\!]$
  $\mathcal{N} \leftarrow [\![\mathcal{G}_\mathcal{Y}; (\mathbf{K}^{(2)})^T, (\mathbf{K}^{(3)})^T]\!]$
  $\mathbf{W} \leftarrow \lceil \mathbf{R}_1/2 \rceil$ left singular vectors of $[\mathbf{M}_{(1)}\ \mathbf{N}_{(1)}]$
  $\mathbf{V} \leftarrow \lfloor \mathbf{R}_1/2 \rfloor$ left singular vectors of $\mathbf{M}_{(1)}$
  $\mathbf{S} \leftarrow \lfloor \mathbf{R}_1/2 \rfloor$ left singular vectors of $\mathbf{N}_{(1)}$
  $\mathcal{G}_\mathcal{X} \leftarrow [\![\mathcal{X}; (\mathbf{W}|\mathbf{V})^T, (\mathbf{U}^{(2)})^T, (\mathbf{U}^{(3)})^T]\!]$
  $\mathcal{G}_\mathcal{Y} \leftarrow [\![\mathcal{Y}; (\mathbf{W}|\mathbf{S})^T, (\mathbf{K}^{(2)})^T, (\mathbf{K}^{(3)})^T]\!]$
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: $RowPixels \times ColumnPixels \times color \times position$. 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.

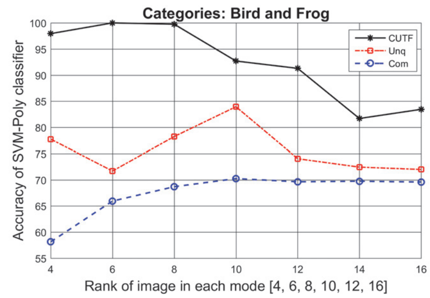

Figure 1: Accuracy comparison by factorization ranks

Categories	Accuracy from LR			Accuracy from SVM-Poly		
	Unq	*Com*	CUTF	*Unq*	*Com*	CUTF
Cat - Dog	.571	.625	**.648**	.535	.787	**.997**
Dog - Frog	.669	.668	**.687**	.675	.858	**.996**
Dog - Deer	.583	.649	**.686**	.655	.841	**.999**
Bird - Frog	.672	.671	**.682**	.687	.782	**.997**
Horse - Dog	.649	.670	**.693**	.704	.808	**.999**
Deer - Horse	.637	.682	**.707**	.678	.843	**.995**
Mobile - Ship	.719	.851	**.864**	.760	.917	**.990**
Mobile - Plane	.737	.777	**.825**	.771	.926	**.985**
Truck - Mobile	.661	.854	**.899**	.685	.945	**.993**
Friedman tests	.0003	.0023	Base	.0003	.0003	Base

Table 1: Comparisons of different factorization methods

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

Acar, E.; Kolda, T. G.; and Dunlavy, D. M. 2011. All-at-once optimization for coupled matrix and tensor factorizations. *arXiv preprint arXiv:1105.3422*.

Bader, B. W.; Kolda, T. G.; et al. 2015. Matlab tensor toolbox version 2.6. Online, http://www.sandia.gov/ tgkolda/TensorToolbox/.

Kolda, T. G., and Bader, B. W. 2009. Tensor decompositions and applications. *SIAM review* 51(3):455–500.

Krizhevsky, A., and Hinton, G. 2009. Learning multiple layers of features from tiny images. *Technical Report, University of Toronto*.

Lahat, D.; Adali, T.; and Jutten, C. 2015. Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of the IEEE* 103(9):1449–1477.

Liu, Y.; Shang, F.; Fan, W.; Cheng, J.; and Cheng, H. 2014. Generalized higher-order orthogonal iteration for tensor decomposition and completion. In *Advances in Neural Information Processing Systems*, 1763–1771.

Ristanoski, G.; Liu, W.; and Bailey, J. 2013. Discrimination aware classification for imbalanced datasets. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, 1529–1532. ACM.