PhD in High-Dimensional Machine Learning with Multiple Measurement Data Vectors

Rémy Boyer and Jérémie Boulanger
University of Lille
CRIStAL-DATING/SIGMA

Keywords : Mathematical optimization, Estimation theory, Machine learning, Tensor Factorisation over Graphs

Increasingly large amount of multidimensional data are being generated on a daily basis in many applications. This leads to a strong demand for learning algorithms to extract useful information from these massive data. Machine learning is the science that deals with the development of algorithms that can learn from data. Typical application is automatic categorization of text into categories, classification between different kinds of data, weather prediction, movie recommendation, spam filtering, ...
Neural Network (NN) is powerful machine learning technique [1, 2]. Typically in the forward pass, two main steps are involved.

1. Linear transformation of the data by a matrix $\mathbf{G}$.
2. Nonlinear activation thanks to a θ-parameter dependent feature map $f_\theta(\cdot) \in \mathbb{R}$ at each layer.

The hidden information associated to the data vector $\mathbf{x}$ is given by

$$f_\theta(\mathbf{G}\mathbf{x}). \tag{1}$$

The second step is the backward pass which is essentially an error backpropagation procedure dedicated to the estimation of the NN parameters, *i.e.*, the matrix $\mathbf{G}$ in different layers based on a score function $h(\mathbf{x})$. Traditionally, an NNs are trained *via* a one-way inputs and the linear transformation is viewed as a one-way factorization. This may constrain the capability of characterizing the structural information in real-world tasks.
This work deals with the extension of NN to multiple measurements. In this case, the trained data are often multidimensional and can be viewed as a multiple measurement vectors (MMV) meaning that the task of interest is the classification of the data matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_N \end{bmatrix}, \tag{2}$$

into one of the categories in set $\mathcal{Y}$. In this scenario, the feature tensor generalizing (1) is naturally defined according to

$$\prod_{j=1}^{N} f_{\theta_{i_j}}(\mathbf{x}_j) \tag{3}$$

and the score function is given by

$$h(\mathbf{X}) = \sum_{i_1,\cdots,i_N} \boldsymbol{\mathcal{G}}(i_1,\cdots,i_N) \prod_{j=1}^{N} f_{\theta_{i_j}}(\mathbf{x}_j). \tag{4}$$

Note that similar cost functions have been considered in [6, 7]. Tensor factorization [9] and deep learning have been rapidly arising as the core technologies in various scientific fields ranging from psychology, chemistry, neuroscience, signal processing, computer vision, and bioinformatics to data mining [8]. Multiple modalities, such as trials, conditions, channels, spaces, times, and frequencies, are ubiquitous in real-world data measurement. Powerful mathematical tools from tensor algebra can be used to extract the salient features from multilinear transformation. Unfortunately, the complexity in term of storage and computational costs to manage the massive tensor $\boldsymbol{\mathcal{G}}$ is combinatorial with the dimensionality/order and this problem is well known under the term of "curse of dimensionality" [3]. In the proposed work, the goal is twofold :

1. explore advanced tensor factorisations [10] for mitigation of the "curse of dimensionality", *i.e.*, to manage the tensor $\boldsymbol{\mathcal{G}}$ with a complexity linear with respect to the dimensionality/order. In particular, factorisations over graph is a promising framework to reach this goal [11]. As illustrated on Fig. 1, a 6-order tensor admits several graph-based representations where the nodes are low-order tensors and the leafs are matrices. The principe here is to break the core tensor $\boldsymbol{\mathcal{G}}$ into a collection 3-order tensors graph-connected : $\{\boldsymbol{\mathcal{G}}_1, \boldsymbol{\mathcal{G}}_2, \boldsymbol{\mathcal{G}}_3, \boldsymbol{\mathcal{G}}_4\}$. Note that the connexion between the leaning model of interest and the multidimensional convolutional Multiple Input Multiple Output model presented for instance in [4, 5].

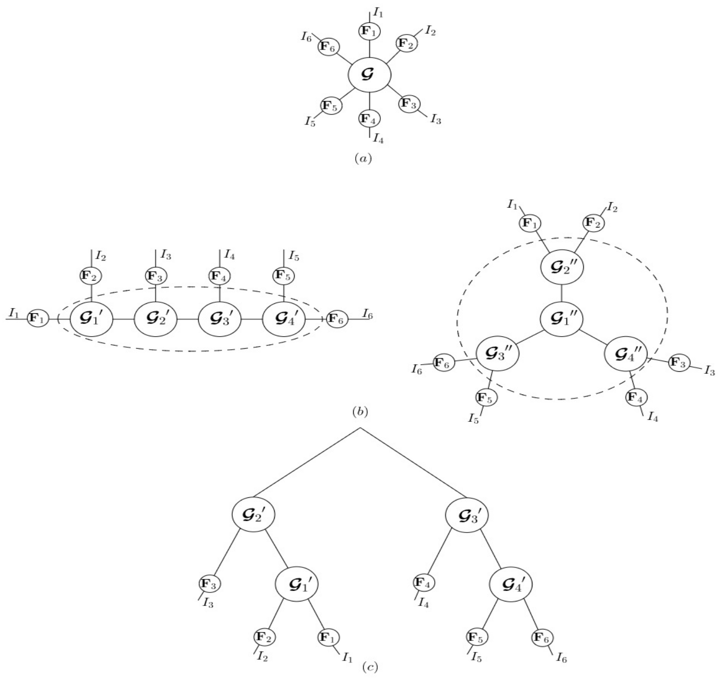

Figure 1 – Graph-based factorization of a 6-order tensor

2. propose new optimisation algorithms for backward pass adapted to the topology of the graph-based tensor factorisation.

Références

[1] Y. LeCun, B. E Boser, J. S Denker, D. Henderson, R. E Howard, W. E Hubbard, and L. D Jackel. Handwritten digit recognition with a back-propagation network. In Advances in neural information processing systems, pp. 396-404, 1990.

[2] Y. LeCun, Y. Bengio, et al. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 3361(10) :1995, 1995.

[3] D.L. Donoho (2000). High-dimensional data analysis : The curses and blessings of dimensionality. AMS Math Challenges Lecture, 1, 32 pages.

[4] R. Boyer, R. Badeau and G. Favier, Fast Orthogonal Decomposition Of Volterra Cubic Kernels Using Oblique Unfolding, IEEE, Proc. of International Conference on Acoustics, Speech, and Signal Processing, (ICASSP'11)

[5] JH. Goulart, M. Boizard, R., Boyer, G., Favier, and P. Comon, Tensor CP Decomposition with Structured Factor Matrices : Algorithms and Performance, IEEE Journal of Selected Topics in Signal Processing, Volume 10, No. 4, June, 2016, pp. 757-769.

[6] N. Cohen, O. Sharir, and A. Shashua. On the expressive power of deep learning : A tensor analysis. In Conference on Learning Theory, pp. 698-728, 2016.

[7] E. Stoudenmire and D. J Schwab. Supervised learning with tensor networks. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (eds.), Advances in Neural Information Processing Systems 29, pp. 4799-4807. Curran Associates, Inc., 2016.

[8] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E.E. Papalexakis, and C. Faloutsos, (2017). Tensor decomposition for signal processing and machine learning. IEEE Transactions on Signal Processing, 65(13), 3551-3582.

[9] A. Cichocki et al., Tensor decompositions for signal processing applications : From two-way to multiway component analysis, IEEE Signal Process. Mag., vol. 32, no. 2, pp. 145-163, Mar. 2015.

[10] I. V. Oseledets, Tensor-Train decomposition, SIAM J. Scientific Computing, vol. 33, no. 5, pp. 2295-2317, 2011.

[11] Y. Znyied, Big Data Processing Using Advanced Tensor Factorizations, PhD L2S/CentraelSupelec, supervision : R. Boyer