Summary of
Semi-supervised Tensor Factorization for Brain Network Analysis

Jiaxin Hu

Email: jhu267@wisc.edu

Date: 2019.2.27

1 Introduction

- Brain network data includes 4 mode: vertices (x2), graphs, time. Each graph is symmetric which means two modes of vertices are symmetric and the 4-mode tensor is partial symmetric. Some graphs are labeled while some are not.

- The goal includes two parts: learning the latent representation(factorization?) from brain network data and semi-supervise classification of task recognition with labeled and unlabeled data.

- **P1** How can we leverage the unlabeled data to facilitate classification?

  Stack all the unlabeled data and labeled data, and do factorization of the data to help the classification of one mode.

- **P2** How can we directly fully utilize the temporal information?

  Consider the temporal information as a mode of the tensor rather than use time-averaged data.

- **P3** How can we effectively fuse representation learning and classification procedures together?

  Blend their training in a unified optimization problem. See the loss function after.

- **P4**How can we achieve feature selection in the latent space?

  Add penalty to the loss to get a fairly sparse result.

2 Preliminaries

This part includes definitions of **Tensor Product, Mode-k Product,Kronecker Product, Khatri-Rao Product, Partially Symmetric Tensor, Mode-k Matricization.**

Another important(I think) definition is rank-one tensor. **An mth-order(mth-mode) tensor is a rank-one tensor if it can be defined as the tensor product of m vectors.**

3 SEMIBAT Framework

3.1 Problem Formulation(Notation)

- $\mathcal{D} = \{G_1, \ldots, G_n\}, |\mathcal{D}| = n$, a dynamic graph dataset of brain network.
- $V, |V| = m$, vertices, all graphs share a parcellation scheme through V, which means the brain is parcellated to m regions. The parcellation of the graphs' rows and columns are symmetric.
- t, the temporal dimensionality.
- G_i, a brain network, can be represented by a partially symmetric tensor $\mathcal{Z}_i \in \mathbb{R}^{m \times m \times t}$
- l, first l graphs are labeled. c, the number of class labels.
- $Y \in \mathbb{R}^{l \times c}$, the class label matrix. $Y_{ij} = 1$ refers to the class that G_i belongs to j-th class, otherwise $Y_{ij} = 0$.

3.2 Tensor Modeling

- Stack all the brain network $\mathcal{Z}_i$ together and we have a 4-ordered (4-mode) tensor $\chi \in \mathbb{R}^{m \times m \times t \times n}$
- **Model(CP factorization)**

$$\chi = \mathcal{C} \times_1 B \times_2 B \times_3 T \times_4 A$$

  where $B \in \mathbb{R}^{m \times k}$ is the factor matrix(in some scenario can be membership matrix),$T \in \mathbb{R}^{t \times k}$ is the factor matrix for time points, ,$A \in \mathbb{R}^{n \times k}$ is the factor matrix for time graphs, and $\mathcal{C} \in \mathbb{R}^{k \times k \times k \times k}$ is a 4-order identity tensor where $\mathcal{C}(i_1, \ldots, i_4) = \delta(i_1 = \cdots = i_4)$.

- To obtain a more concise and interpretable result, add an orthogonality constriants $A^T A = I$.

- **Interpretation of Factorization**

  The process of CP factorization decompose 4-mode tensor into k rank-one 4-mode tensor. In this case, the $f, f = 1, 2 \ldots, k$ rank-one 4-mode tensor can be represented by the tensor product of 4 vectors, *i.e.* $B_{:,f} \circ B_{:,f} \circ T_{:,f} \circ A_{:,f}$.

  Therefore, each element in the 4-mode tensor χ can be decompose to k parts, for example:

  $$\chi_{(3,4,1,2)} = B_{31}B_{41}T_{11}A_{21} + \cdots + B_{3k}B_{4k}T_{1k}A_{2k}$$

  where in the model identity matrix $\mathcal{C}$ used to guarantee that the tensor can be split by separated k rank-one tensors.

- **My Extension** In this paper, we want to use factorization to describe the tensor as much as possible and the element in the tensor can be regard as continuous variables. Therefore, the factorization process looks like some feature extraction method like PCA in lower dimension.

  If we consider a clustering scenario, that needs to cluster different region of the tensor with 4 modes together, how should we transfer the model?

  The model can also write in the same factorization formula and noise:

  $$\chi = \mathcal{C} \times_1 B \times_2 B \times_3 T \times_4 A + \epsilon$$

  Consider change the identity tensor $\mathcal{C}$ to a tensor $\mathcal{U} \in \mathbb{R}^{k_1 \times k_2 \times k_3 \times k_4}$ in which each element refers to the mean of each cluster μ_{xyzq}. And k_i means the $i - th$ mode can be cluster as k_i groups. At this time, B, T, A would be membership matrix with binary elements $\{0, 1\}$, for instance, $T_{ij} = 1$ refers to the elements with index i in the third mode belongs to the j-th group on the third mode, otherwise $T_{ij} = 0$. As to the noise, assume each noise $\epsilon_{i_1,i_2,i_3,i_4} \sim_{i.i.d} N(0, \sigma^2)$. At this time:

  $$\mathbb{E}\chi_{(3,4,1,2)} = \mu_{xyzq}B_{3x}B_{4y}T_{1z}A_{2q} = \mu_{xyzq},\ assume\ \chi_{(3,4,1,2)} \in Group_{xyzq}$$

  If we relax the condition of binary elements in some modes, for instance, relax the binary condition of matrix $A \in \mathbb{R}^{n \times p}$ and $\mathcal{U} \in \mathbb{R}^{k_1 \times k_2 \times k_3 \times p}$.And keep the B, T as membership matrix.

Then we have:

$$\mathbb{E}\chi_{(3,4,1,2)} = \mu_{xyz1} B_{3x} B_{4y} T_{1z} A_{21} + \cdots + \mu_{xyzp} B_{3x} B_{4y} T_{1z} A_{2p}$$
$$assume\ \chi_{(3,4,1,2)} \in Group_{xyz}\ on\ the\ mode\ B, B, T$$

- **Classification by regression** Come back to the problem of this paper, assume a matrix of regression matrix $W \in \mathbb{R}^{k\times c}$ which assigns the graphs with labels based on the graph factor matrix A. Then we have a linear regression model on mode A:

$$Y = DAW + noise,\ D = [I^{l\times l}, 0^{l\times(n-l)}] \in \mathbb{R}^{l\times n}$$

- **Loss** Consider a loss function emerge factorization error , classification loss(regression loss) and the penalty of the classification parameters. This loss incorporate the two procedures together and use the labeled data and unlabeled data together. The function is :

$$\min_{B,T,A,W} \|\chi - \mathcal{C} \times_1 B \times_2 B \times_3 T \times_4 A\|_F^2 + \alpha\|DAW - Y\|_F^2 + \lambda\|W^T\|_{2,1}$$
$$s.t. A^T A = I$$

where α, λ are positive tuning parameters.

- **Optimization Framework** The key idea is to optimize the objective with respect to one variable while fixing others and decouple constraints using an Alternating Direction Method of Multipliers (ADMM) scheme.

- **Time Complexity** Overall, the updates of all model parameters require $O(k^3 + (m+n+t)k^2 + (m^2nt + n^2)k + n^3)$ arithmetic operations in total.