CASSANDRA (draft)
A framework for tensorial prediction
v0.1

January 18, 2018

1 Abstract

We present a method for tensorial prediction using a convolutional blind decomposition approach (filters convolved with activations) where the activation variables are shared between the input and the output.

2 A framework for tensor evolution prediction using a tensor-based spectral convolutional method

2.1 Analysis Formulation - Multiple Tensors + Spectral Data fit

Let $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_D]$ be a collection of D tensors.

> **Analysis formulation:** Identification of a common compositional structure (filters) among the D tensors such that the D tensors can be reconstructed, within a margin of error, by a finite sum of tensorial convolutions of the compositional structures (filters) and an activations of the compositional structures (activations).

This can be achieved through the following optimization problem:

$$f, \begin{bmatrix} h_1 \\ \vdots \\ h_D \end{bmatrix} = \arg\min_{f,h} \left\| \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_D \end{bmatrix} - \sum_k f_k * \begin{bmatrix} h_{1,k} \\ \vdots \\ h_{D,k} \end{bmatrix} \right\|, \qquad (1)$$

where f_j denotes the jth compositional structure (filter) and $h_{i,j}$ denotes the activation of the jth compositional structure on the ith tensor (X_i). This problem is biconvex and, generally, ill-defined, as the solution might not be unique (multiple combinations of filters and activations can reconstruct the collection of tensors). To address the biconvexity of the problem, we find an approximate solution through an alternating optimization of the filters and their activations, in convex subproblems (addressed further ahead). To address the fact that this optimization problem can be ill-defined, we can introduce priors on the activations (*e.g.* sparse activations, activations smooth or piecewise smooth activations in specific dimensions, ...) or in the filters (*e.g.* filters with fixed components in a number of dimensions, smooth or piecewise smooth activations in specific dimensions, ...). Our problem becomes then:

$$f, \begin{bmatrix} h_1 \\ \vdots \\ h_D \end{bmatrix} = \arg\min_{f,h} \underbrace{\left\| \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_D \end{bmatrix} - \sum_k f_k * \begin{bmatrix} h_{1,k} \\ \vdots \\ h_{D,k} \end{bmatrix} \right\|}_{\text{data fit}} + \underbrace{\mathcal{P}_h(h) + \mathcal{P}_f(f)}_{\text{prior information}}, \quad (2)$$

where $\mathcal{P}_h(h)$ denotes a prior on the activations and $\mathcal{P}_f(f)$ denotes a prior on the filters.

The solution of the data fit component of the optimization problem is computationally challenging as the data and the filters increase in number of dimensions and size. We note however, that we can efficiently solve the data

fit problem in the spectral domain as

$$\arg\min_{f,h} \left\| \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_D \end{bmatrix} - \sum_k f_k * \begin{bmatrix} h_{1,k} \\ \vdots \\ h_{D,k} \end{bmatrix} \right\| =$$

$$\arg\min_{f,h} \left\| \mathcal{F}\left(\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_D \end{bmatrix} - \sum_k f_k * \begin{bmatrix} h_{1,k} \\ \vdots \\ h_{D,k} \end{bmatrix} \right) \right\| =$$

$$\arg\min_{f,h} \left\| \mathcal{F}\left(\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_D \end{bmatrix} \right) - \mathcal{F}\left(\sum_k f_k * \begin{bmatrix} h_{1,k} \\ \vdots \\ h_{D,k} \end{bmatrix} \right) \right\| =$$

$$\arg\min_{f,h} \left\| \mathcal{F}\left(\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_D \end{bmatrix} \right) - \sum_k \mathcal{F}(f_k)\mathcal{F}\left(\begin{bmatrix} h_{1,k} \\ \vdots \\ h_{D,k} \end{bmatrix} \right) \right\| \tag{3}$$

where $\mathcal{F}$ denotes the Fourier transform.

2.2   Prediction Formulation

Let $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_D]$ be a collection of D tensors representing D different social signals, and $\mathbf{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_M]$ a collection of M tensors representing M target social signals (social signals we want to predict). Our goal is the estimation of $\mathbf{Y}$ from the knowledge of the *directed* patterns of interaction $\mathbf{X} \to \mathbf{Y}$.

The estimation of the directed patterns of interaction allows not only to understand patterns of interaction from a causality point of view (not approached in this document) but also to estimate $\mathbf{Y}$ given $\mathbf{X}$. This is of utmost importance in the prediction of time-series and tensor evolution.

We approach the problem from the same approach as the analysis part.

$$f_1, h_1 = \arg\min_{f,h} \|\mathbf{X} - \sum_{k=1}^{k_1} f_k * h_k\| + \mathcal{P}_h(h) + \mathcal{P}_f(f), \tag{4}$$

where $\mathcal{P}_h$ and $\mathcal{P}_f$ denote priors on the activations and filters, respectively. This problem can be solved in the spectral domain as follows,

$$f_1, h_1 = \arg\min_{f,h} \|\mathbf{X} - \sum_{k=1}^{k_1} \mathcal{F}(f_k)\mathcal{F}(h_k)\| + \mathcal{P}_h(h) + \mathcal{P}_f(f), \tag{5}$$

3

where $\mathcal{F}(.)$ denotes the Fourier transform operator. Assuming convexity of the priors on the activations and filters, this is a biconvex problem, which can be solved approximately as,

$$f^{t+1} = \arg\min_{f} \|\mathcal{F}(\mathbf{X}) - \sum_{k=1}^{k_1} \mathcal{F}(f_k)\mathcal{F}(h_k^t)\| + \mathcal{P}_f(f), \tag{6}$$

$$h^{t+1} = \arg\min_{h} \|\mathcal{F}(\mathbf{X}) - \sum_{k=1}^{k_1} \mathcal{F}(f_k^{t+1})\mathcal{F}(h_k)\| + \mathcal{P}_h(h). \tag{7}$$

This corresponds to finding simultaneously the filters and the activations for $\mathbf{X}$.

We then find the filters g such that,

$$g = \arg\min_{g} \|\mathcal{F}(\mathbf{Y}) - \sum_{k=1}^{k_1} \mathcal{F}(g_k)\mathcal{F}(h_k)\| + \mathcal{P}_h(h) + \mathcal{P}_g(g). \tag{8}$$

This means that we find filters g such that we can reconstruct $\mathbf{Y}$ from the filters g and the activations derived from $\mathbf{X}$.

The pair of filters f and g provide a path between the known and accessible $\mathbf{X}$ and the unknown and $\mathbf{Y}$,

train

$$(\mathbf{X}_{\text{train}}) \rightarrow (\underline{f}, h) \tag{9}$$
$$(\mathbf{Y}_{\text{train}}, h) \rightarrow (\underline{g}) \tag{10}$$

predict

$$(\mathbf{X}_{\text{test}}, \underline{f}) \rightarrow (h_{\text{test}}) \tag{11}$$
$$(h_{\text{test}}, \underline{g}) \rightarrow (\mathbf{Y}_{\text{predicted}}) \tag{12}$$