针对 EEG 信号数据的张量分解

摘要

脑电图学(EEG)是脑功能图成像的基础工具。EEG 信号通常被组织成向量或者矩阵的形式，这样组织数据有利于进行时序分析、谱分析和矩阵分解等。事实上，EEG 信号分为时间、空间等不同类型，他们可以表示为一个多路的矩阵，称为张量。这篇综述从三个方面介绍了张量分解在 EEG 信号研究中的进展。首先，介绍了当前现有的模态和 EEG 张量表示。其次，介绍两种张量分解算法并且对这两种算法做了对比研究。最后介绍了两种分解算法在 EEG 信号中的应用。论文特别介绍了两种分解算法中组成个数的选择问题。文章最后介绍了 N 路偏最小二乘和高阶偏最小二乘，这有助于解决不同模态的脑电信号。

1. 介绍

数据分析处理在基于脑成像的研究中发挥着重要作用。记录的脑电成像数据可以表示为一维序列、二维矩阵或者多维张量。不同的数据表示形式采用不同的分析方法。

1.1 一维和二维数据

例如，在脑电图学的早期研究阶段，EEG 数据通常表示为一个时间序列，所有的数据都表示为一个向量。那个阶段，能量谱分析方法用来研究 EEG 时序数据中的振荡行为。接着，时序分析里面的时频分析方法被用来分析 EEG 信号。现在，研究使用多路电极来收集 EEG 数据。因此，EEG 数据至少包含了时间和空间维度的信息。自然而然地，EEG 数据就表示成矩阵形式。一些专业的 EEG 分析软件会将 EEG 数据显示在屏幕上，横轴代表时间维度，纵轴代表空间信息。因此，二维信号分析方法，例如主成分分析和独立成分分析被用来提取感兴趣的脑活动事件。

1.2 EEG 数据的多路特性

在 EEG 实验中，除了时间和空间信号还可能存在其他类型的数据。例如，EEG 分析中，经常需要比较不同组之间的回应（健康控制组和诊断组的不同回应）。因此，这种情况下就多出现了一个数据类型——组别。又如，在探寻事件内在相关性的试验中，实验变量包括不同组和不同的刺激环境。以上例子说明 EEG 数据是包含了多种类型的多路阵列。

然而大多数脑研究计算工具都是针对一维或者二维数据。相应地，为了方便二路信号数据处理，除了时间和空间之外的信号通常串联到一起（数据水平地连到一起），或者将他们组织成栈式结构（垂直地连到一起）。这种方法是将多路阵列信号伸展成矩阵的形式。对于 EEG 数据而言，这无疑会损失折叠类型之间的信息，例如时间、频率和空间类型。这些交互信息可能是研究的兴趣所在。因此，适用于多路通道的数据处理方法能准确揭示多通道数据之间的相互作用。

1.3 多路阵列是张量，这是一种表示和处理数据的新方法

多路阵列数据称作张量。对于矩阵而言，矩阵分解经常被用作数据处理和分析。对于张量而言，张量分解也具备同样的作用。张量分解能够揭示多种模式的内在相互作用。张量首先在数学中定义，而后被用到心理测绘学的多路数据分析中。现有的经典综述介绍了张量分解的历史、模式、算法及各种应用。

最近，张量分解在信号处理中大放光彩。事实上，从上世纪 80 年代以来，张量分解就被用来分析 ERP 数据。最近十年，有越来越多的报告介绍张量分解在 EEG 数据中的使用。然而，还没有文章专门介绍张量分解在 EEG 信号分析中的应用。这篇文章总结了张量分解在 EEG 信号分析中的应用，并探讨了与此相关的核心问题。

文章开始介绍 EEG 数据的张量表示以及张量分解模型。接着，文章介绍了如何将张量分解应用到 EEG 数据分析处理中，最后文章预测了张量分解在 EEG 信号处理中的可能趋势。许多文章介绍了张量分解相对于矩阵分解的优势，这里就不赘述。这篇文章主要介绍张量分解的基本概念及其应用。

2. EEG 数据多模态和高阶张量

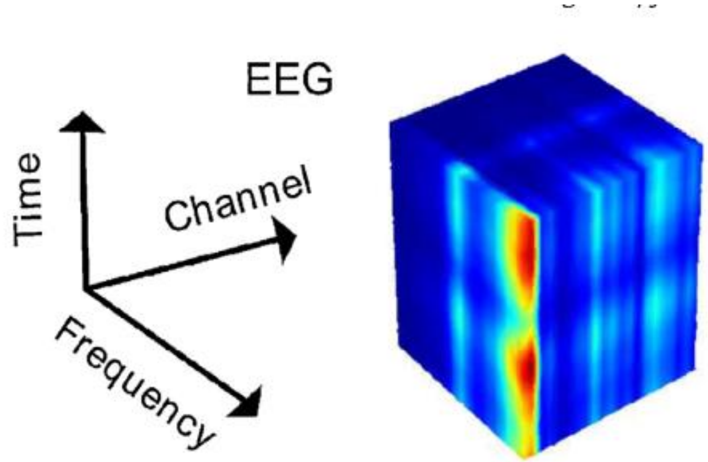

Figure 1

EEG 数据可以分为三种类型：自发 EEG 数据，ERPs 数据和正在进行的 EEG 数据。无论是哪种类型的数据，原始数据都是几分钟或者几小时的连续数据。EEG 数据通常根据刺激条件分割。除了时间和空间，附加的分割类型也是一类数据。如果 EEG 数据转化到频域，将会多出频率类型。对于 ERP 数据而言，实验过程中会有各种实验条件，这将会产生更多的实验数据类型。对于同一实验项目，属于不用组的人会产生组别这一新的实验数据类型。对于不同的实验项目，不同的实验项目也可以当作一个新的类型。

因此，EEG 实验至少存在 7 种信号，包括时间、频率、空间、试验、条件、项目和组别。对于传统的 ERP 研究，试验类型在 EEG 信号平均之后就会消失。高阶张量保存了实验中真实存在的七种信号。显然这个张量可能非常巨大。

当张量的维度超过三维后，就无法直观地可视化数据。图一是一个三阶张量，

包含了时间、频率和空间三种类型的数据。这个三阶张量由多通道 EEG 数据的时频表示。他同时展示了脑活动的时间、频谱和空间演化过程。如果将三维信号串联或者栈式组成二维矩阵将很难有上述效果。

时间序列表示的 EEG 信号可以转换成其他类型的数据，转换之后这些信号就可以用波形之外的形式表示。例如，如果功谱图用来分析 EEG 时序数据，那么这些数据就表示成不同频率分量上的强度。频率分量的强度可以称作是一种特征或者表示。不同类型的 EEG 数据已经被用来做神经障碍诊断。因此，当 EEG 数据转化为其他类型的表示时就多了一种特征类型。

3. 张量分解模型

CP 和 Tucker 分解是张量分解的两种基本模型。CP 分解又可以称为并行特征分析或者标准分解。

3.1 标准多阶分解

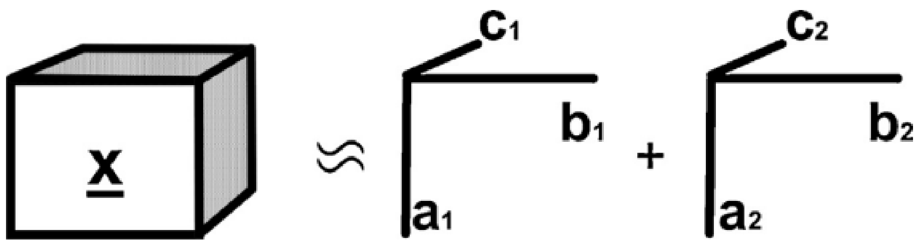

Figure 2

3.1.1 定义

给定一个三阶张量，CPD 分解如下：

$$X = a_1 \circ b_1 \circ c_1 + a_2 \circ b_2 \circ c_2 + E \approx X_1 + X_2 \quad (1.1)$$

图一是三阶张量的表示。二分量 CP 分解之后，可以提取出两个时序、两个谱、两个空间分量。这个分解中，时序分量 a_1 、谱分量 b_1 和空间分量 c_1 彼此相关，他们的外积组成了秩 1 张量 X_1 。第二个时间、谱和空间分量的外积组成了秩 1 的张量 X_2。由上可知，CP 分解就是将原张量分解成一连串的秩 1 张量之和加上一个误差张量。

广义地看，给定一个 N 阶张量，CP 分解具有以下形式：

$$X = \sum_{r=1}^{R} u_r^{(1)} \circ \cdots \circ u_r^{(N)} + E = \sum_{r=1}^{R} X_r + E = \mathrm{X} + E \approx X \quad (1.2)$$

对于张量分解而言，根据模式的属性可以给各种模式添加不同的约束。例如，如果要求数据是非负的，那么就可以使用非负约束。

3.1.2 CP 和矩阵分解的关系

如果 N=2，公式 1.2 退化成矩阵分解

$$X=\sum_{r=1}^{R}u_r^{(1)}\circ u_r^{(2)}+E=X+E \quad (1.3)$$

公式 1.3 是盲源分离模型。因此，我们可以说式 1.3 是二源分离，式 1.1 是关于 N 个模型的分离。如果没有进一步的约束，式 1.3 没有确切的解。这就是矩阵分解和 CP 分解的最大区别，只有给定一个约束条件，张量分解才会有确定的解。

3.2 Tucker 分解

3.2.1 定义

$$X=\sum_{r_1=1}^{R_1}\cdots\sum_{r_N=1}^{R_N}g_{r_1\cdots r_N}a_{r_1}^{(1)}\circ\cdots\circ a_{r_N}^{(N)}+E \quad (1.4)$$

其中 $X_{r_1\cdots r_N}=a_{r_1}^{(1)}\circ\cdots\circ a_{r_N}^{(N)}, I_n\ge R_n$， $g_{r_1\cdots r_N}$ 组成了核心张量 G。为了避免不必要的误会，Tucker 使用 a 代替 CP 分解里的 u。因此，Tucker 分解是是 $R_1\times\cdots\times R_N$ 和误差张量的和的形式。

理论上，Tucker 分解没有特定的解。实际中，当为不同模式添加不同的约束后，Tucker 分解可以得到确定解。

3.2.2 CP 分解和 Tucker 分解的区别

对于两种基本的张量分解模型，CP 分解和 Tucker 分解之间有四点主要的区别。

在 CP 分解中，每个模式的组成部分个数是固定的。然而，Tucker 分解中，不同模式的组成部分的个数是不同的。

CP 分解和 Tucker 分解都可以表示成秩 1 张量之和的形式，每个张量又是由 N 的组成的外积组成。对于 CP 分解而言，每个组成分量需要满足一定的约束条件，而对 Tucker 分解而言，对于组成矩阵内的组成分量没有约束。例如，对于上述时序分量 a_1 、谱分量 b_1 和空间分量 c_1，他们三个彼此相关，但是他们和 a_2、b_2 和 c_2 没有联系。但在 Tucker 分解中，不同组成成分矩阵中的组成可能彼此相关。CP 分解中的核心矩阵是一个等同矩阵，而 Tucker 分解中的核心张量可能是任意合适大小的张量。理论上，只要施加一定的约束，CP 分解就有确定解，而不对 Tucker 分解施加额外的约束，便无法获得确定解。

4. 张量分解在 EEG 信号处理中的应用

将张量分解应用到 EEG 信号处理中，目的是为了定位脑活动，以研究神经认知和神经诊断问题。在这个研究中，有两个问题需要考虑。第一个问题是需要使用 CP 分解还是 Tucker 分解。第二个问题是张量分解是在单个个体层面还是在群体层面。考虑之前的分析，张量通常包含了一个 EEG 实验里单一项目的某一分段的数据。而对于后者，张量通常包含了同一项目或者不同项目的多组数据。因此，本篇综述从上述层面来对张量分解的应用进行分类。进一步，我们需要考虑张量分解中组成的个数如何选择。

下文中，我们将使用样本来描述 EEG 实验里的一次试验或一个项目。

4.1 CP 分解应用在单个样本中

如在图一中显示的，EEG 数据的 TRF 组成了一个三阶张量。这样的张量可以使用 CP 分解。

相应的，空间分量可以被用来做源定位。时序和谱分量同时反应了数据的时序和谱特征，这个结果可以用来做病例诊断。

4.2 CP 分解应用在多样本张量中

单个样本的张量分解被用来做可视化检查。对于不同的样本，如果张量分解时独立进行的，那么很难从整体上对数据进行分析。这是因为张量分解固有的不确定性。因此，在多通道的张量上来研究张量分解时十分必要的。这包括特征提取、特征选择和特征分析。这个过程被广泛应用于模式识别和机器学习中。

4.2.1 特征提取

当张量包含了样本这一模式，提取的分量包含了每个模式的表示。通常，样本模式被安排在张量的最后一个模式。给定 N 阶张量 $X \in R^{I_1 \times \cdots \times I_N}$ ，I_N 是样本模式中所有样本的个数。CP 分解可以写作

$$X = \sum_{r=1}^{R} u_r^{(1)} \circ \cdots \circ u_r^{(N-1)} \circ f_r + E = I \times U_1^{(1)} \times \cdots \times U_{N-1}^{N-1} \times F_N + E$$

其中，对于 $n = 1, \cdots,$ N-1 ，有 $\| U_r^{(n)} \|_2 = 1$ ， $U_r^{(n)} \in R^{I_N \times 1}$ 表示模式 N 里面的第 r 个分量，因此，如果样本模式包含了一个项目的多种条件或多个组别，我们可以对特征组成 f_r 提取出的特征进行分析。

例如，对于一个包含时间、频率、空间和样本模式的四阶张量，通过一个施加非负约束的 CP 分解可以分成四个组成分量：

$$X = I \times U_1^{(t)} \times U_2^{(s)} \times U_3^{(c)} \times F_4 + E = \sum_{r=1}^{R} u_r^{(t)} \circ u_r^{(s)} \circ u_r^{(c)} \circ f_r + E$$

其中 $U^{(t)}$ 代表时序矩阵，$U^{(s)}$ 代表谱分量矩阵，$U^{(c)}$ 代表空间分量矩阵；F 代表

了多空间的特征分量矩阵。需要注意的是，上式所有的元素都是非负的。

图三显示了利用 CP 分解对四阶张量分解的结果。每个模式的前三个分量都显示在图中。张量的维度是 71*60*9*42。每个模式提取出来的分量个数是 36。图三的每一行代表了一个特征的时序、谱和空间分量。特别需要注意的是，图三中每行的分量和特征和其他行的分量和特征无关。这是 CP 分解相对于 Tucker 分解的独特之处。

4.2.2 特征选择和分析

当一个张量被 CP 分解后，下面就需要考虑如何选择和分析提取出来的特征。通常特征分析的目的决定了特征选择方法。众所周知，EEG 信号包含了感兴趣的大脑活动、不感兴趣的大脑活动和噪声数据。因此通过张量分解提取的特征依旧包含了感兴趣或者不感兴趣的特征。我们希望提取出感兴趣的脑活动的特征，用以做进一步的特征分析。

总的来说，针对不同的研究问题，有两种数据分析方法。一种是统计方法，例如变量分析，被用来验证之前所做的假设。另一种方法是模式识别里面的分类器，它用来分类已经确定的假设。对于前者我们可以使用 P 个值来代表不同因素的重要性。对于后者而言，使用百分比来表示预测、诊断和分类的确定程度。因而，对于前者来说，特征选择通常是基于 EEG 信号的属性，用不到项目的组标记。后者通常使用自适应的特征选择算法，依赖于项目的组标签。

对于后者，所有的样本通常被划分成训练集和测试集。首先，训练集的部分被分解，并通过自适应的特征选择算法选择特征来训练分类器。在成功训练后，新样本数据被投射到训练集分量上来获取新的特征。使用训练好的分类器可以获得新样本的标签。成功训练和测试后，分类器可以用来做特征研究。

实际上，两种类型对应着不同的研究问题。前者研究为什么，后者研究怎么做。前者经常出现在神经科学领域而后者经常出现在计算机科学领域。在一定程度上，这两个问题可以结合到一起。

4.3 Tucker 分解应用到多样本张量

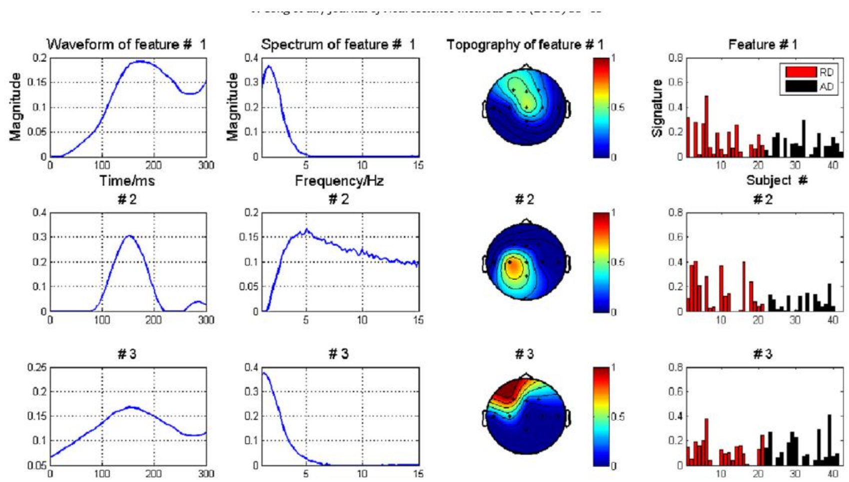

Figure 3

Tucker 分解经常应用到各种脑活动成像数据中。这个应用包括特征选择、特征提取和特征分析。

当样本模式被安排在 N 阶张量的最后，应用 Tucker 分解可以通过样本模式的分量或者核心张量提取特征。接下来的特征选择和特征分析也可以像 CP 分解那样分解成两类。

图四展示了添加了非负约束的 Tucker 分解处理四阶张量的结果。分解后有 8 个时序分量，四个谱分量和 6 个空间分量。核心张量 G 提取了特征，一共 192 个特征。正如上文提到的那样，CP 分解提取出的第二个分量是感兴趣的分量，可以用来做深入研究。有意思的是第八个时序分量、第三个谱分量和第四个空间分量和图三中第二行对应的分量是非相似。需要注意的是，Tucker 分解提取出的分量和其他模式里的分量是相关的。这是 CP 分解和 Tucker 分解的核心区别。

依据上文提到的四个区别，Tucker 分解比 CP 分解更为复杂。Tucker 分解可以使用更少的分量来提取张量里感兴趣的特征。如果想要使用 EEG 信号的先验知识来选择感兴趣的特征，当使用 Tucker 分解时就需要使用到每个模式里的这些信息。而对于 CP 分解而言，这不是必须的。

4.4 每个模式里组成个数的确定

当使用张量分解时，一个重要的问题就是确定参数，例如每个模式提取分量个数的确定。对于 CP 分解而言，只需要确定一个参数即可。而对于 Tucker 分解而言，参数的个数和模式的个数一样多。确定参数的方法包括 DIFFT、ARD 和模型阶数选择。

众所周知，矩阵分解和张量分解都可以用来做盲源分离。因此为了确定 CP 分解提取分量的个数，最好考虑一下，矩阵分解分量的个数。独立成分分析是一种矩阵分解算法，被广泛地应用到 EEG 信号处理中。当使用 ICA 时，提取分量的个数被认为是大脑中信号源的数量。无论是 EEG 数据还是其他大脑成像数据，一个源包括了空间分量和时间分量，他们的外积产生了一个源的模式。因此，源的个数实际上是源模式的个数。在大脑功能成像数据中，源的个数可以有几十个。当使用 CP 分解来处理 EEG 数据时，ICA 估计可以用来近似决定提取分量的个数。

然而，使用交叉验证的方法和基于 ARD 的贝叶斯学习只认为需要提取几个分量，这与 EEG 数据的实际特征不相符。

模型阶数选择从理论上讲是与张量分解中源个数的确定相匹配。然而，当信噪比较低时，他们无法产生精确的结果。实际上，EEG 信号，特别是 EEG 数据，通常是信噪比较低。

令人惊奇地是，对 CP 分解使用 DIFFT 产生了和矩阵分解相近的分量个数。DIFFT 是根据 CP 模型或者 Tucker 模型分量个数的增长而改变的。下面是一个 CP 分解的例子：

$$fit(m) = 1 - \frac{\| X - X_m \|_F}{\| X \|_F}$$

其中，X_m 是对原始数据 X 的一个近似表示，$\|\bullet\|_F$ 是 Frobenius 范数，$fit(m)$

是单调递增的。两个相邻的 fit 之间的差如下：

$$dif(m) = fit(m) - fit(m-1)$$

相邻 fit 之间的比率如下：

$$DIFFT(m) = \frac{dif(m)}{dif(m+1)}$$

当 DIFFT 最大时，就认为 CP 分解是原始数据的最好近似。图五显示了 DIFFT 在 CP 分解中的应用。

DIFFT 的缺点是非常耗时。因此，当线下进行张量分解时，DIFFT 是一个合适的选择。

5. 潜在的趋势

近来，并发 EEG 数据 fMRI 受到广泛关注，他们主要被用来克服 EEG 数据低空间分辨率的缺点。因此，如何处理两种形态的数据变成了重要的研究问题。当前处理两种形态数据的方法大多是矩阵分解。然而，当 EEG 数据和 fMRI 数据用张量表示时，我们可以使用张量分解来处理数据。

实际上，只要一种模式的数据使用张量表示，我们就可以使用张量分解算法来提取特征。针对两种形式数据的张量分解算法又可以称为 N 路偏最小二乘和高阶偏最小二乘。

5.1 标准偏最小二乘

对于脑研究而言，偏最小二乘用来寻找大脑数据 X 和行为数据 Y 之间的共同隐向量。这些向量能够表明大脑和行为数据之间的协方差。标准 PLS 如下：

$$X = TP^T + E_X$$

$$Y = UC^T + E_Y$$

其中，T 包含了从大脑数据中提取出的正交向量；U 是行为数据提取出的正交向量。在偏最小二乘框架下，U 和 T 有最大的协方差，表示 Y 的最简单的模型如下：

$$U = TD$$

其中。D 是一个对角阵，$d_{rr} = u_r^T t_r / t_r^T t_r$ 。

5.2 N 路偏最小二乘

当大脑数据表示成张量 X 的形式，N 路偏最小二乘可以用来代替标准偏最小二乘。例如，当大脑数据用四阶张量表示，有项目、时间、空间和刺激类型四种形式的数据。当使用 N 路 PLS 时，CP 模型可以将张量 X 分解成 R 个分量，可以表示如下：

$$X = \sum_{r=1}^{R} t_r \circ p_r \circ q_r \circ s_r + E_X$$

$$Y=\sum_{r=1}^{R} d_{rr} t_r c_r^T + E_Y$$

其中t_r 是隐向量，他的元素和 I 个项目相对应。 p_r, q_r, s_r 是载入向量。他们有时被称为因素。在 N 路 PLS 下，他们会满足如下条件：

$$\{p,q,s,c\} = \arg\max_{p,q,s,c} [\mathrm{cov}(t,u)]$$

$$s.t. t_i = \sum_{l=1}^{L} \sum_{k=1}^{K} \sum_{j=1}^{J} x_{ilkj} p_l q_k s_j = YC$$

$$and \parallel p \parallel_2^2 = \parallel q \parallel_2^2 = \parallel s \parallel_2^2 = 1$$

5.3　　高阶偏最小二乘

考虑张量 X，Tucker 模型可以表示如下：

$$X=\sum_{r=1}^{R} G_r \times t_1^r \times P_2^r \times Q_3^r \times S_4^r + E_{X,R}$$

$$Y=\sum_{r=1}^{R} d_{rr} t_r c_r^T + E_{Y,R}$$

其中，R 是隐向量个数，t_r 是第 r 个隐向量，P_r, Q_r, S_r是对应的载入向量，G_r是第 r 个核心张量，反映了第 r 个隐向量和对应载入向量之间的交互关系。

尽管只有少量的研究使用张量分解来处理分析 EEG 信号和 fMRI 信号,但是，从理论上讲，N 路偏最小二乘和高阶偏最小二乘是研究两种模式数据的有力工具。实际应用中，当计算大量数据时，需要使用超级计算机来进行运算。

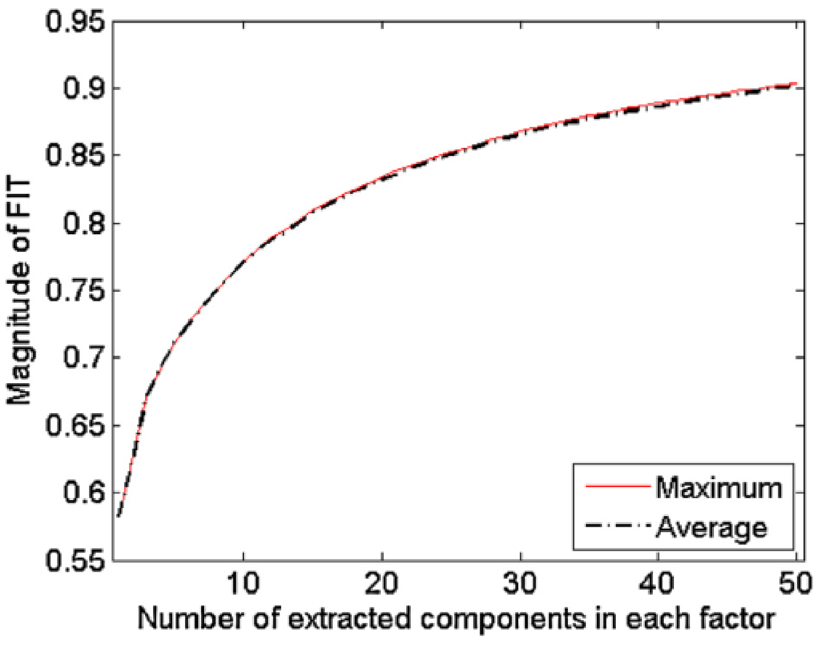

Figure 4