ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 32,No. 2,Mar. 2017,pp. 266－277
DOI:10.16337/j.1004-9037.2017.02.006

http://sjcj.nuaa.edu.cn
E-mail:sjcj@nuaa.edu.cn
Tel/Fax:＋86-025-84892742

非负组合模型及其在声源分离中的应用

张雄伟[1]　李轶南[1]　时文华[1,2]　胡永刚[1]　陈栩杉[3]

(1.解放军理工大学指挥信息系统学院,南京,210007;2.空军航空大学教官基地,蚌埠,233000;3.武警政治学院政工信息化教研室,上海,201703)

摘　要:非负组合模型在人工智能、数据挖掘和智能信息处理研究领域具有十分重要的应用意义,已经逐渐成为声源分离中最常使用以及最具代表性的模型之一。内含于其中的非负成分的加性组合与人类听觉系统的感知机理高度契合。利用非负组合模型进行声源分离的技术正在变得越来越流行。本文从被称作非负矩阵分解的最基本的非负组合模型开始,首先回顾了非负组合模型的基本原则,包括需要求解的基本问题、目标函数的度量以及求解相关问题的常用方法。在此基础上,系统地讨论了非负矩阵分解在声源分离不同应用领域的拓展。最后指出并讨论非负组合模型研究中有待进一步研究的开放问题。

关键词:声源分离;非负组合模型;非负矩阵分解

中图分类号:TN912.3　　**文献标志码**:A

Non-negative Compositional Models and Its Application in Acoustic Source Separation

Zhang Xiongwei[1], Li Yinan[1], Shi Wenhua[1,2], Hu Yonggang[1], Chen Xushan[3]

(1. College of Command Information System, PLA University of Science and Technology, Nanjing, 210007, China; 2. Flight Instructor Training Base, Air Force Aviation University, Bengbu, 233000, China; 3. Lab of Political Information, People's Armed Police Institute of Politics, Shanghai, 201703, China)

Abstract: Non-negative compositional models are of great importance in the application of artificial intelligence, data mining and intelligent information processing research. They have gradually become one of the most representative and frequently used models of acoustic source separation in recent years. The embedded additive combination of non-negative components matches well with the characteristic of human perception. Techniques that make use of non-negative compositional models have been increasingly popular in acoustic source separation. Starting from the most basic non-negative compositional model, which is termed as non-negative matrix factorization (NMF), we firstly review the principles of non-negative compositional model, including the basic problem to be solved, the measurement of objective function and some typical methods to solve related problems. Based on these principles, we systematically discuss the variety extensions of NMF designed for particular applications in acoustic source separation. Finally, some open problems are presented and discussed.

Key words: acoustic source separation; non-negative compositional model; non-negative matrix factorization

---

基金项目:国家自然科学基金(61471394,61402519)资助项目;江苏省自然科学基金(BK20140071, BK20140074)资助项目。
收稿日期:2015-06-05;**修订日期**:2016-06-30

引 言

如何准确、高效地发掘和表示样本数据中潜在的特征是机器学习和数据挖掘领域的一个重要问题。实际应用中的很多数据可以视作由一些“部分”通过“组合”的结合方式所构成的。这种组合的构成方式,意味着构成部分之间不会出现相减或抵消的现象,而是通过叠加的方式相互组合而成。典型的组合数据包括人口、自然语言处理中字词出现频度等计数类数据。学者们提出了大量的数学模型来表示和处理此类数据[1],旨在与这类数据的组合叠加本质相吻合。这些数学模型的核心思想是通过非负元素的非负组合来实现对于数据整体的表示和描述,进而更好地把握和挖掘出观测数据整体内含的基本规律。本文将以“非负组合模型”来指代具备上述特征的数学模型。非负组合模型可以大致分为两类:(1)以矩阵分解为代表的线性叠加组合模型;(2)在以神经网络为代表的非线性模型上添加非负性,从而构建非线性的叠加模型。尽管非负组合模型最适合于处理计数数据,但是近年来,基于非负组合模型的各类方法也被广泛应用于语言习得[2]、音乐分析[3]、图像分类[4]、波束成型[5]和高光谱分析[6]等诸多领域。在处理多声源信号分离这样的老问题中,也取得了比较理想的效果。这主要是由于两点原因:(1) 非负组合模型所得到的结果具有物理上的可解释性,因此能够较容易地与客观物理现象找到合理的对应关系(很多物理信号,不可能存在负的构成分量),无论是各个频率分量的大小,还是某个特定频率成分出现的时间及其相应的强度变化,都能在非负组合模型下得到很好的表示和解释。(2) 这种组合叠加的方式与人类感知系统通过对于客观事物从部分到整体的认知过程相契合。例如,在听交响乐时,人类听觉系统的感知过程,可以被认为是通过将听到的弦乐器、管乐器和打击乐器等各种乐器的分别感知来实现对于交响乐整体的感知。

在人工智能和机器学习领域,为了提高所设计计算系统的智能水平,常常借鉴人类大脑对于信息的处理机制。在此方面,非负组合模型由于其简洁(纯加性的方式容易导致稀疏的表示结果,使得对于数据的描述变得较为简洁)、灵活(能够很容易地与涌现出的新方法相融合)的特点,其合乎人类大脑感知的直观体验,具有确定物理意义的计算结果[7-8],展现出强大的生命力。本文将从最基本的非负组合模型,即非负矩阵分解开始,逐步阐述非负组合模型的基本定义和求解方法;然后结合声源分离的实际应用,对各类基于非负组合模型的算法进行介绍和综述。最后总结全文,指出有待进一步解决的开放问题。

1 非负矩阵分解

非负组合模型是一种非常灵活的数据表示模型,能够很方便地与最新的诸如深度神经网络、稀疏字典学习算法及稀疏低秩分解等相融合并拓展出新的基于非负组合模型的算法。这一系列算法全都起源于基于乘法迭代的非负矩阵分解(Nonnegative matrix factorization, NMF)[9]。

1.1 非负矩阵分解的定义及基本问题

给定一个非负的数据矩阵 $\boldsymbol{V}\in\mathbf{R}_{\geqslant 0}^{F\times N}$(即维数为 $F\times N$ 的非负矩阵),非负矩阵分解问题旨在实现的分解为

$$\boldsymbol{V}\approx\hat{\boldsymbol{V}}=\boldsymbol{W}\boldsymbol{H} \tag{1}$$

式中:$\boldsymbol{W}$ 和 $\boldsymbol{H}$ 分别为大小为 $F\times K$ 和 $K\times N$ 的非负矩阵。由于该分解通常用于对非负数据的逼近和降维,因此,K 通常以 $(F+N)K\ll FN$ 的方式来选取,确保对于数据规模的压缩。此外,$\boldsymbol{W}\boldsymbol{H}$ 通常可以看作是对于 $\boldsymbol{V}$ 的低秩逼近,记为 $\hat{\boldsymbol{V}}=\boldsymbol{W}\boldsymbol{H}$ 。

式(1)的分解结果通常使用如下的最小化问题来求解,即有

$$\min_{\boldsymbol{W},\boldsymbol{H}} D(\boldsymbol{V}\mid\boldsymbol{W}\boldsymbol{H})\quad \text{s.t. } \boldsymbol{W}\geqslant 0,\ \boldsymbol{H}\geqslant 0 \tag{2}$$

式中:记法 $\boldsymbol{A}\geqslant 0$ 表示矩阵 $\boldsymbol{A}$ 的所有元素都非负。其中,$D(\boldsymbol{V}\mid\boldsymbol{W}\boldsymbol{H})$ 是一个可分离的测量,即

$$D(\boldsymbol{V} \mid \boldsymbol{WH}) = \sum_{f=1}^{F} \sum_{n=1}^{N} d([\boldsymbol{V}]_{fn} \mid [\boldsymbol{WH}]_{fn}) \tag{3}$$

其中 $d(x \mid y)$ 是一个标量的目标函数。目标函数是在给定 x 的条件下关于 y 的正值函数，只有在 $x=y$ 时取得最小值。

1.2 目标函数的刻画与度量

为了求解式(3)中的优化问题，首先确定目标函数，用以刻画重构数据对观测数据的逼近程度。在 NMF 中，最常见的是以 β 散度作为度量标准定义的一系列目标函数。β 散度最早由 Basu 等提出[10]，这一系列目标函数本质上讲是一些关于参数 β 的函数，不同的 β 取值对应着不同的噪声假设，其定义为

$$d_{\beta}(x \mid y) = \begin{cases} \dfrac{1}{\beta(\beta-1)}(x^{\beta} + (\beta-1)\, y^{\beta} - \beta x y^{\beta-1}) & \beta \in \mathbf{R} \backslash \{0,1\} \\ x \log \dfrac{x}{y} - x + y & \beta = 1 \\ \dfrac{x}{y} - \log \dfrac{x}{y} - 1 & \beta = 0 \end{cases} \tag{4}$$

最常见的 β 取值主要有 $\beta=2$，$\beta=1$ 和 $\beta=0$，分别对应欧式距离，广义 KL 散度(Generalized Kullback-Leibler, KL divergence)以及板仓散度(Itakura-Saito, IS divergence)。该参数从本质上来说，是对观测噪声的统计学特性进行假设。例如当 β 分别取 2，1 和 0 时，相应的观测噪声则分别对应加性高斯噪声、泊松噪声和乘性伽马噪声。其关系如表 1 所示(表中 $\hat{\boldsymbol{v}}_n = \boldsymbol{W}\boldsymbol{h}_n$，其所对应的系数为 $\hat{v}_{fn}$)。

β 散度可以看做是关于 β 的连续函数，图 1 给出了式(4)在 $x=1$ 条件下，不同取值 β 下的对应的关于变量 y 的变化曲线(图中的所有值均无量纲)。需要指出的是，对于任意 β 的取值，都满足式(5)所呈现出的尺度特性[11]，该性质可以表示为

$$d_{\beta}(\lambda x \mid \lambda y) = \lambda^{\beta} d_{\beta}(x \mid y) \tag{5}$$

这表明当 $\beta > 0$ 时，左右目标函数的主要是那些数值较大(能量较大)的取值；而 $\beta < 0$ 则恰恰相反，是那些比较小的值。当 $\beta=0$ 时所对应的 IS 散度恰好具备尺度不变性，这是其他 β 取值下所不具备的性质。

表 1 常见的散度及其对应的概率生成模型

Tab. 1 Common divergences and their corresponding generative models

散度 $D(\boldsymbol{v}_n \mid \hat{\boldsymbol{v}}_n)$	内含的生成模型 $p(\boldsymbol{v}_n \mid \hat{\boldsymbol{v}}_n)$
欧式距离	Additive Gaussian
$\frac{1}{2\sigma^2}\sum_f (v_{fn} - \hat{v}_{fn})^2$	$\prod_f N(v_{fn} \mid \hat{v}_{fn}, \sigma^2)$
广义 KL 散度	Poisson
$\sum_f \left(v_{fn} \log \frac{v_{fn}}{\hat{v}_{fn}} - v_{fn} + \hat{v}_{fn}\right)$	$\prod_f P(v_{fn} \mid \hat{v}_{fn})$
IS 散度	Multiplicative Gamma
$\sum_f \left(\frac{v}{\hat{v}_{fi}} - \log \frac{v_{fi}}{\hat{v}_{fi}} - 1\right)$	$\prod_f G\left(v_{fn} \mid \alpha, \frac{\alpha}{\hat{v}_{fn}}\right)$

1.3 非负矩阵分解的求解算法

NMF 算法交替更新 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 两个矩阵，首先给定 $\boldsymbol{W}$ 更新 $\boldsymbol{H}$，然后给定 $\boldsymbol{H}$ 更新 $\boldsymbol{W}$。由于 NMF 的对称性，以上两个步骤在本质上是相同的($\boldsymbol{V} \approx \boldsymbol{WH}$ 等同于 $\boldsymbol{V}^{\mathrm{T}} \approx \boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}$，而 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 的位置正好发生了互换)。因此，聚焦于如下的子问题，则

$$\arg\min_{\boldsymbol{H}} D(\boldsymbol{V} \mid \boldsymbol{WH}) \quad \text{s.t. } \boldsymbol{H} \geqslant 0 \tag{6}$$

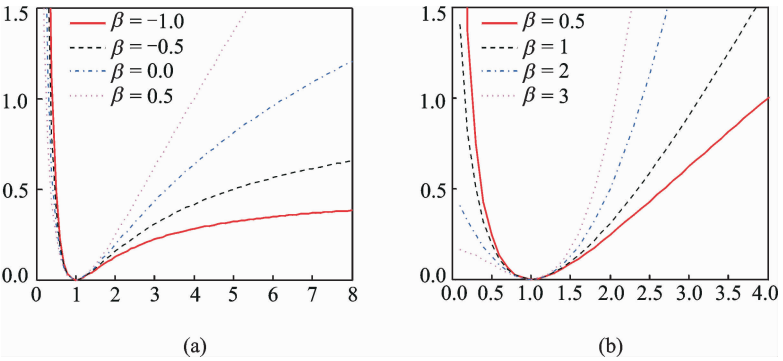

图 1 在 NMF 中使用的典型散度函数说明

Fig. 1 Illustration of typical divergence functions used in NMF

式中固定 $\boldsymbol{W}$,准则函数 $C(\boldsymbol{H})$可以划分为 $\sum_n D(\boldsymbol{v}_n \mid \boldsymbol{W}\boldsymbol{h})$,其中 $\boldsymbol{v}_n$ 和 $\boldsymbol{h}_n$ 分别是 $\boldsymbol{V}$ 和 $\boldsymbol{H}$ 的第 n 列。这样,需要求解的问题可以转化为

$$\arg\min_{\boldsymbol{H}} D(\boldsymbol{v} \mid \boldsymbol{W}\boldsymbol{h}) \quad \text{s.t.} \quad \boldsymbol{h} \geqslant 0 \tag{7}$$

式中 $\boldsymbol{v} \in \mathbf{R}_+^F$,$\boldsymbol{W} \in \mathbf{R}_+^{F\times K}$ 和 $\boldsymbol{h} \in \mathbf{R}_+^K$。

(a) 人脸数据集测试样本 (b) NMF分解所得结果

(a) Samples from datasets of faces (b) Result of NMF

图 2 非负矩阵分解在人脸数据集上的运行结果

Fig. 2 Result of NMF on human faces dataset

求解 NMF 问题最直观的方法是采用传统的梯度下降法,最早提出的正矩阵求解就是采用这样方法来进行求解[12],然而加法迭代的过程中难以避免地会产生负值,一般需要将这些负值强制置零。文献[9,13]巧妙地采用了乘法迭代的方式来获得 NMF 的更新公式,避免了出现强制置零的现象,这种采用乘法迭代实现对 NMF 进行求解的方法已经成为求解此类问题的经典方法。此外,这种只做加法不做减法的计算方式,得到的计算结果往往具有稀疏特性,使得对于数据的表示更加清晰简洁。图 2 给出了在人脸数据集上运行前后的结果,可以看出,分解所得的非负基函数,与人脸的局部信息(例如鼻子、眉毛和眼睛等)相类似,因此实现了对人脸从“部分”到“整体”的表示。

常用的用于获得乘法迭代更新公式的途径有两种:一种是通过构造辅助函数进行优化而得到更新公式的优化最小(Majorization minimization,MM)算法;另一种则是基于变化步长的启发式乘法算法(Heuristic multiplicative algorithm)。优化最小算法具有严格的理论基础,会保证推导出的 NMF 最终收敛到一个局部最小值,缺点是寻找并构造满足要求的辅助函数比较困难,文献[11]给出了各种 β 取值

下的辅助函数，本文不再赘述；启发式乘法算法缺乏理论支撑，其本质上是变化步长的梯度下降算法，但推导简便，并且一般情况下也能获得令人满意的结果，因此也被经常使用。

假设 $\nabla_{\boldsymbol{h}}D(\boldsymbol{h})$ 是 $D(\boldsymbol{h})$ 关于 $\boldsymbol{h}$ 的偏导数，则 $\nabla_{\boldsymbol{h}}D(\boldsymbol{h})$ 可以拆分为 $\nabla_{\boldsymbol{h}}D(\boldsymbol{h})=\nabla_{\boldsymbol{h}}^{+}D(\boldsymbol{h})-\nabla_{\boldsymbol{h}}^{-}D(\boldsymbol{h})$ 的形式（其中，正部 $\nabla_{\boldsymbol{h}}^{+}D(\boldsymbol{h})$ 和负部 $\nabla_{\boldsymbol{h}}^{-}D(\boldsymbol{h})$ 均非负）。若将梯度下降的步长设定为 $\eta=\dfrac{\boldsymbol{h}}{\nabla_{\boldsymbol{h}}^{+}D(\boldsymbol{h})}$，就能得到相应的乘法迭代公式，推导过程为

$$\boldsymbol{h}^{(t+1)}\leftarrow\boldsymbol{h}^{(t)}-\eta\nabla_{\boldsymbol{h}^{(t)}}D(\boldsymbol{h}^{(t)})=\boldsymbol{h}^{(t)}-\frac{\boldsymbol{h}^{(t)}}{\nabla_{\boldsymbol{h}^{(t)}}^{+}D(\boldsymbol{h}^{(t)})}\left[\nabla_{\boldsymbol{h}^{(t)}}^{+}D(\boldsymbol{h}^{(t)})-\nabla_{\boldsymbol{h}^{(t)}}^{-}D(\boldsymbol{h}^{(t)})\right]=$$

$$\boldsymbol{h}^{(t)}-\boldsymbol{h}^{(t)}+\frac{\nabla_{\boldsymbol{h}^{(t)}}^{-}D(\boldsymbol{h}^{(t)})}{\nabla_{\boldsymbol{h}^{(t)}}^{+}D(\boldsymbol{h}^{(t)})}\boldsymbol{h}^{(t)}=\boldsymbol{h}^{(t)}\frac{\nabla_{\boldsymbol{h}^{(t)}}^{-}D(\boldsymbol{h}^{(t)})}{\nabla_{\boldsymbol{h}^{(t)}}^{+}D(\boldsymbol{h}^{(t)})} \tag{8}$$

除了经典的乘法迭代算法之外，近年来还涌现出很多其他的快速求解方法，例如采用基于交替乘子法（Alternating direction method of multipliers, ADMM）的更新方法[14]和 Nesterov 最优梯度[15]等方法。从统计视角来看，上述方法本质上在不同的概率模型下，使用期望最大算法来进行最大似然估计，进行参数估计。文献[16]在此基础上，进一步考虑了更加复杂的推断方法（包括变分贝叶斯（Variational Bayesian, VB）以及马尔科夫链蒙特卡洛（Markov chain Monte Carlo, MCMC））来计算边际似然，进而进一步提升了模型的表示能力。然而，为了计算边际似然的积分，VB 方法常常需要挑选满足共轭性质的先验概率分布来使计算变得简洁；MCMC 方法则由于其过于巨大的计算开销限制了其在实际工程中的应用。

2 声源分离中的非负组合模型

基于非负组合模型的声源分离算法大体上可以分为 2 类：一类处理的是多声源分离问题，采用的技术手段主要包括全监督模式、半监督模式以及无监督模式这 3 种声源分离模式；另一类处理的是去混响问题，研究如何从密闭空间中，与反射信号相混合的混合声音信号中恢复出纯净的原始声源信号的问题。在上述两类 4 种分离条件下，通常假设混合声源的幅度谱由不同声源的幅度谱相互叠加而成，尽管现实世界中的信号并不是如此简单的叠加，但是这样的假设通常能够得到令人满意的结果[17]。

2.1 全监督模式下的声源分离

全监督模式下，具备最多的先验信息，各个声源样本均可以提前获得，一般采用非负组合模型来对样本信号进行表示[18-20]。以语噪分离为例，在全监督模式下，已知语音和噪声的样本，可以预先使用适当的非负组合模型来获得语音的基矩阵 $\boldsymbol{W}_{\mathrm{S}}$ 和噪声的基矩阵 $\boldsymbol{W}_{\mathrm{N}}$。然后，通过将语音基矩阵和噪声基矩阵连结起来，构成带噪语音信号的联合字典 $\boldsymbol{W}=[\boldsymbol{W}_{\mathrm{S}},\boldsymbol{W}_{\mathrm{N}}]$。随后，固定联合字典，将带噪语音信号在联合字典上进行投影，得到的分解结果为

$$\boldsymbol{v}=[\boldsymbol{W}_{\mathrm{S}},\boldsymbol{W}_{\mathrm{N}}]\begin{bmatrix}\boldsymbol{h}_{\mathrm{S}}\\ \boldsymbol{h}_{\mathrm{N}}\end{bmatrix} \tag{9}$$

最后，通过维纳滤波器估计出纯净语音信号的幅度谱为

$$\hat{\boldsymbol{s}}=\frac{\boldsymbol{W}_{\mathrm{S}}\boldsymbol{h}_{\mathrm{S}}}{\boldsymbol{W}_{\mathrm{S}}\boldsymbol{h}_{\mathrm{S}}+\boldsymbol{W}_{\mathrm{N}}\boldsymbol{h}_{\mathrm{N}}}\odot\boldsymbol{v} \tag{10}$$

其中，除法是矩阵对应位置的元素相除，符号“⊙”则表示对应元素之间的乘法。

在全监督条件下，基于非负组合模型的语噪分离可以用图 3 来表示。文献[19]通过添加稀疏约束，在一定程度上提升了语音噪

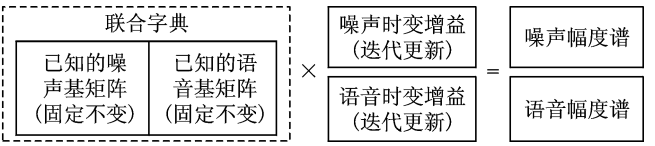

图 3 全监督模式下的更新方式

Fig. 3 Update strategy of supervised fashion

声分离的效果。在文献[20]中,Paris 基于类似的方法,实现了男女说话人的有监督分离,此种方法需要语音和噪声(或者男女说话人)的样本作为先验信息,然而在实际应用中,并不一定能够具备如此多的先验信息来进行预先训练。半监督和无监督分离算法正是用于解决先验信息不足问题的方法。

2.2 半监督模式下的声源分离

半监督模式所具备的先验信息少于全监督模式,在进行声源分离前,只有一部分(而非全部)声源已知。例如在语噪分离中,特定背景噪声的声音样本可以在无语音的间歇期获得,然而具体说话人的声音却无法预先获得,这种条件下的声源分离被称为半监督模式下的声源分离[21-24]。在半监督模式下,固定已知的声源训练所得的非负基矩阵不变,未知声源的基矩阵通常采用随机初始化的方式进行初始化,迭代地更新剩余的 3 个矩阵(包括未知声源的基矩阵、已知声源的时变增益和未知声源的是时变增益),即可进行分离(所需先验信息少于全监督模式,但多于后面将要介绍的无监督模式)。

上述过程如图 4 所示,在图 4 中,假设噪声信号为已知信号而语音信号为未知信号。图 5 给出了半监督模式下,基于非负组合模型的时频字典学习语噪分离算法[22],其思路与图 4 一致,独特之处在于采用了卷积非负矩阵分解(Convolutive non-negative matrix factorization, CNMF)来对噪声信号进行建模,从而能够更好地建模出噪声信号内部所存在的短时相关性。CNMF 对于声学信号短时变化的建模,是通过在 NMF 的基础上,引入一个额外的时间维度来实现的。此时,式(1)的就被重新改写为

$$\boldsymbol{V} \approx \hat{\boldsymbol{V}} = \sum_{t=0}^{T-1} \boldsymbol{W}(t) \overrightarrow{\boldsymbol{H}}^{t} \tag{11}$$

式中:$\boldsymbol{V}$ 为需要进行分解的非负数据,$\hat{\boldsymbol{V}}$ 为 $\boldsymbol{V}$ 的逼近表示;$\boldsymbol{W}(t) \in \mathbf{R}_{\geqslant 0}^{F \times K}$,$\forall t$ 为一系列基函数,每一个 t 值选取对应着一个时间切片,通过这样的 T 个时间切片,即可实现对音频信号短时相关性的建模;$\boldsymbol{H}$ 为对应的增益矩阵;符号 $\overrightarrow{(\cdot)}^{t}$ 表示右移矩阵的列,与之对应的,$\overleftarrow{(\cdot)}^{t}$ 则表示左移矩阵的列,如下给出了关于上述符号的直观解释,即

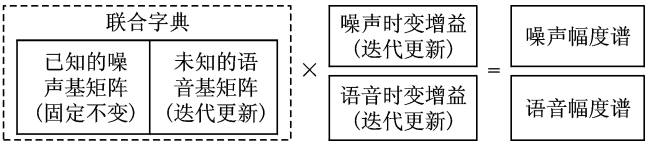

图 4 半监督模式下的更新方式

Fig. 4 Update strategy of semi-supervised fashion

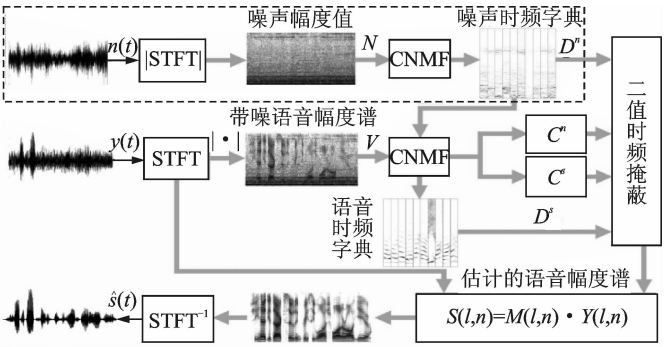

图 5 时频字典学习语噪分离算法框图

Fig. 5 Block diagram of speech and noise separation based on the time-frequency dictionary learning

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} \quad \overrightarrow{\boldsymbol{A}}^{1} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 0 & 5 & 6 & 7 \end{bmatrix}$$

$$\overleftarrow{\boldsymbol{A}}^{1} = \begin{bmatrix} 2 & 3 & 4 & 0 \\ 6 & 7 & 8 & 0 \end{bmatrix} \quad \overrightarrow{\boldsymbol{A}}^{2} = \begin{bmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 5 & 6 \end{bmatrix}$$

2.3 无监督模式下的声源分离

在资源较少的条件下进行无监督模式的声源分离,常常只有少量的先验信息可以利用。以语音和噪声分离为例,在某些应用场景中,语音和噪声特征并不可提前预知,只能根据语音和噪声的固有特点进行建模和分离[25]。相较于有监督模式(包括全监督和半监督模式),无监督模式具备较少的先验信息,因此,通常情况下获得的结果会劣于有监督模式。但是其不依赖先验知识的特性,使得无监督模式

也有非常广泛的应用。基于非负组合模型的无监督模式下的语噪分离主要有3类实现方法：

(1) 通过语音断点检测(Voice activity detection, VAD)来区分带噪语音和纯噪声[26-27]。图6为文献[27]中给出的无监督语音增强算法，在该算法中，通过VAD来判断语音间歇期，并在语音间歇期使用在线的卷积非负稀疏编码算法(Convolutive non-negative sparse coding, CNSC)来学习不断更新噪声的基函数，进而实现语音增强。

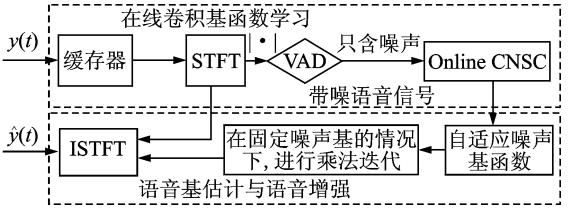

图6 在语音间歇期在线学习噪声基函数的无监督算法

Fig. 6 Unsupervised algorithm which learns the noise bases in an online fashion during the pauses of speech

(2) 无监督分离算法，基于通用语音字典的增强方法，即训练一个能够囊括所有说话人特征的通用语音字典，来实现将目标语音信号从背景噪声中分离出来的目的[28-30]。Sun和Mysore分别对每个说话人构建一个小的非负基矩阵(局部字典)。然后将这些小的非负基矩阵连结起来构成一个较大的语音字典(全局字典)，在使用时，通过添加组稀疏约束，一次只激活其中的一个或几个局部字典，很好地实现了无监督下语音和噪声的分离[28]。Kim等在文献[28]的基础上，提出了能够更好保持声源流形的(Mixture of local dictionaries, MLD)的方法[29]，进一步提高了增强算法的效果。Germain等在文献[28]的基础上，开发出了能够实现在线处理的增强方法[30]。

(3) 将非负组合模型与稀疏低秩分解(Sparse and low-rank decomposition)相融合来实现语噪分离[31-34]。在某种特定的噪声环境中，背景噪声的时频结构会随着时间的推移反复出现，呈现出一定的低秩结构；与之相对应的，语音信号由于富有变化、具有谐波结构等固有特征，会表现出稀疏结构。因此，采用稀疏低秩模型即能够实现无监督模式下的语音噪声分离[34]。Huang首次将在基于稀疏低秩结构的鲁棒主成分分析(Robust principal component analysis, RPCA)应用到音乐伴奏分离，并取得了很好的分离效果[35]。然而，RPCA并不能保证分离结果的非负性，需要对非负部分进行强制置零。为了解决此问题，Sun等将非负组合模型与稀疏低秩分解想融合，即保持了稀疏低秩模型在无监督条件下分解的优势，又使得分解结果具备了物理上的可解释性，通过信息融合的方式进一步提升了增强效果[31]。Li等在此基础上进一步发展，考虑到了背景噪声特定模式在时间上的重复性，利用自相关方法确定重复模式的长度，引入卷积基函数来对重复模式进行建模，并推导了相应的迭代公式，从而提出了新的增强方法[32]，所用算法的流程如图7所示。

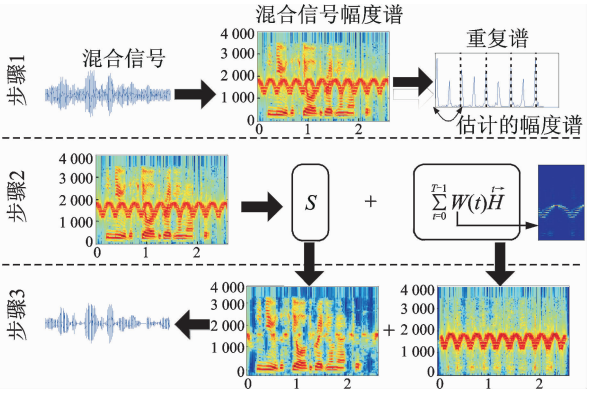

图7 考虑背景噪声短时连续性的无监督增强算法

Fig. 7 Unsupervised speech enhancement considering temporal continuity of background noise

2.4 声源信号的去混响问题

声源去混响是另一类比较特殊的声源分离问题。从本质上讲，声源只有一个，但是由于在密闭空间内，存在不可预知的多径问题，导致由于多径效应反射回来的音频信号与原始信号相叠加，造成原始信号质量下降[36]。近年来，非负组合模型也被尝试用于解决去语音混响问题。文献[37]基于非负卷积变换函数(Non-negative convolutive transfer function, N-CTF)和NMF模型提出了去混响算法。

假设 $s(n)$ 和 $h(n)$ 分别是离散时间纯净语音信号和M接头(M-tap)的室内冲击响应(Room impulse response, RIR)(M-tap RIR还可以看作是一个长度为 M 的滤波器)。由于声音的多径效应所混叠

的语音信号可以使用纯净语音信号幅度谱 $[S[f,t]]$ 和 M 接头滤波器的幅度相应的卷积来获得[38],即

$$|Y[f,t]|\approx\sum_{\tau=0}^{M}|S[f,t-\tau]||H[f,\tau]|\equiv|S[f,t]|*|H[f,t]| \tag{12}$$

对于纯净信号和 RIR 滤波器的盲估计是不可行的,因为存在模型模糊问题[1]。为此需要为模型引入适当的先验信息,例如使用基于 NMF 的语音字典[39]或是引入稀疏性约束[40]。因此,实际应用中常常使用另外的非负组合模型来对 $|S[f,t]|$ 进行表示。

3 非负组合模型的在线处理与阶数选择

在使用非负组合模型处理声学信号时,常常需要面对两个问题:(1) 在线处理。因为音频信号是典型的流式数据,在一些应用中,无法完全采集下来再进行处理,而是需要不断地根据信号的变化,保持对于基函数的不断更新,即实现在线处理;(2) 阶数的选择。在对声源信号进行建模时,究竟需要多少个基函数来实现对于特定声源的表示,是阶数选择问题研究的核心问题。过少的基函数将无法承载丰富的声源信息,而过多的基函数不仅会浪费宝贵计算资源和内存,更容易产生过拟合的问题,同样影响最终的分离效果。因此选取刚好够用的基函数非常重要。

3.1 非负组合模型的在线处理

在线处理是一种针对海量或流式数据的处理方式。在进行在线处理时,通常需要设定一个统计量,用于存储过去样本的信息,当新样本到来时,通过一定的方式将其内化转变为内部统计量的信息,最终实现在线处理[41]。式(13~14)给出欧式距离下 CNSC 的迭代公式为

$$\boldsymbol{H}\leftarrow\boldsymbol{H}\odot\frac{[\boldsymbol{W}(p)]^{\mathrm{T}}\overleftarrow{\boldsymbol{V}}^{p}}{[\boldsymbol{W}(p)]^{\mathrm{T}}\overleftarrow{\hat{\boldsymbol{V}}}^{p}+\lambda\cdot\boldsymbol{\Xi}} \tag{13}$$

$$\boldsymbol{W}(p)\leftarrow\boldsymbol{W}(p)\odot\frac{\boldsymbol{V}\overrightarrow{\boldsymbol{H}}^{p\mathrm{T}}}{\hat{\boldsymbol{V}}\overrightarrow{\boldsymbol{H}}^{p\mathrm{T}}} \tag{14}$$

式中:$\boldsymbol{\Xi}$ 表示一个与 $\boldsymbol{V}$ 同样大小的全 1 矩阵,而 λ 是用于控制编码矩阵稀疏程度的参数。

Wang 等[42]在此基础上,通过对式(14)交换次序,引入统计量 A 和 B,得到了 CNSC 的欧式距离下的在线处理公式,即

$$W(p)\leftarrow W(p)\odot\frac{\sum_u\dot{B}(p,u)}{\sum_q W(q)\sum_u\dot{A}(q,p,u)} \tag{15}$$

其中训练数据被切分为等长的小块,u 为训练样本切片的序号(此处假设各个切片之间相互独立),并且

$$\dot{A}(q,p,u)=\overrightarrow{\boldsymbol{H}}_u^{q}\overrightarrow{\boldsymbol{H}}_u^{p\mathrm{T}} \tag{16}$$

$$\dot{B}(p,u)=\boldsymbol{V}_u\overrightarrow{\boldsymbol{H}}_u^{p\mathrm{T}} \tag{17}$$

分别为样本切片 u 的对于统计量 A 和 B 的贡献。因此,前面 u 个样本切片所对应的统计量为

$$A(q,p;u)=\sum_{t=1}^{u}\dot{A}(q,p,t) \tag{18}$$

$$B(p;u)=\sum_{t=1}^{u}\dot{B}(p,t) \tag{19}$$

伴随着新样本的不断到来以及式(15)的反复迭代,基函数根据新样本的特性不断地对原有基函数进行调整和改变,而过去样本的统计信息则通过 A 和 B 不断积累和存储起来。通过上述方式,不断地在原有样本的基础上添加新样本信息的处理方法即为在线处理。文献[27]则推导出了在广义 Kull-

back-Leibler 散度下，CNSC 的在线学习公式，并将其应用于无监督语音增强中。Lefèvre 给出了在 Itakura-Saito 散度下在线学习的更新公式[43]。

3.2 非负组合模型的阶数选择

当训练样本量规模一定时，如何选择合适的阶数（用于表示非负观测数据的非负基函数的个数），就变成一个非常重要的问题。进行阶数选择的方法有很多，组稀疏[44]、计算密集的马尔科夫链蒙特卡罗（Markov chain Monte Carlo, MCMC）[45]、正则化熵[46]和自相关决策（Automatic relevance determination, ARD）[47]等方法均被用于进行阶数选择。

其中文献[47]结合 ARD，引入将 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 相连结的超参数 λ（假设 W_{fk} 和 H_{kn} 服从参数为 λ_k 的半高斯分布或指数分布，如式(20,21)所示），并通过最大后验概率（Maximum a posteriori, MAP）估计的方法自动获得合理的阶数。

(1) 当为 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 分配半高斯分布时有

$$\begin{cases} p(W_{fk} \mid \lambda_k) = HN(W_{fk} \mid \lambda_k) \\ p(H_{kn} \mid \lambda_k) = HN(H_{kn} \mid \lambda_k) \end{cases} \tag{20}$$

式中：当 $x \geqslant 0$ 时，$HN(x \mid \lambda) \triangleq \left(\frac{2}{\pi\lambda}\right)^{\frac{1}{2}} \exp\left(-\frac{x^2}{2\lambda}\right)$；当 $x < 0$ 时，有 $HN(x \mid \lambda) = 0$。

(2) 当为 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 分配指数先验时，有

$$\begin{cases} p(W_{fk} \mid \lambda_k) = \varepsilon(W_{fk} \mid \lambda_k) \\ p(H_{kn} \mid \lambda_k) = \varepsilon(H_{kn} \mid \lambda_k) \end{cases} \tag{21}$$

式中：当 $x \geqslant 0$ 时，$\varepsilon(x \mid \lambda) \triangleq \frac{1}{\lambda}\exp\left(-\frac{x}{\lambda}\right)$；当 $x < 0$ 时，则有 $\varepsilon(x \mid \lambda) = 0$。

对于每个超参数 λ_k，$\forall k$，添加 Inverse-Gamma 先验，即

$$p(\lambda_k; a, b) = IG(\lambda_k \mid a, b) = \frac{b^a}{\Gamma(a)}\lambda_k^{-(a+1)} \exp\left(-\frac{b}{\lambda_k}\right) \tag{22}$$

式中：a 和 b 分别为非负的形状和尺度超参数。所有的 λ_k，$\forall k$ 共享相同的形状和尺度超参数。文献[48]进一步将其拓展为张量形式，实现了对于具备一定时频结构模式阶数的自动选择。文献[49]使用变分贝叶斯期望最大（Variational Bayes expectation maximization, VBEM）的方法，通过对增益矩阵的各个变量进行积分，改进了文献[47]中的算法，通过计算出充分统计量，解决了所需要估计随样本个数增加而线性增大的问题，然而采用变分方法为了计算方面的便利，往往需要选择一些具备共轭性质的分布对。这样的分布虽然便于进行变分求解，然而所做的先验并不总是与实际情况很好吻合。Hoffman 基于贝叶斯非参数化方法提出了伽马过程非负矩阵分解（Gamma Process Nonnegative Matrix Factorization, GaP-NMF）算法，在一定程度上使得假设更趋合理[50]。

4 结束语

本文综述了非负组合模型及其在声源分离中的应用。从非负组合模型的定义和优化问题的建立出发，讨论了目标函数的构造与度量，给出了启发式求解算法的推导，作为接下来具体应用的铺垫。在基于非负组合模型的声源分离领域，分全监督、半监督和无监督 3 种情况分别予以系统的分析、总结和比较。此外，对于非负组合模型经常遇到的在线处理和阶数选择问题，亦给与了介绍和总结，以期为了解非负组合模型算法研究的现状和相关工作的开展提供有益参考。尽管目前基于非负组合模型的声源分离算法取得了一定的成果，然而目前依然存在很多开放的问题有待进一步解决，这些问题主要包括：(1) 求解非负组合模型的高效计算方法。(2) 如何在具有噪声污染或数据缺失的自然数据中自适应地选取出所需要的阶数。(3) 在有监督声源分离时，训练数据与实际数据之间难免会出现一些出入，当偏差相

对较大时,分离算法的性能将会显著下降,如何设计出一种能够依据实际样本特征来进行自适应调整和改变的模式(包括说话人自适应调整和背景噪声自适应学习),依然是有待进一步研究的问题。(4)现有的非负组合模型,往往是基于相对简单的线性叠加方式来进行运算,这样的方法虽然简单易行,然而在面对海量数据时,并不能很好地刻画出数据内在的规律,文献[24]将非负组合模型和深度神经网络的基本思想相结合,利用神经网络的非线性特性,产生了一些初步的研究成果,然而相关领域的发展依然方兴未艾。(5)时序信号的相关性建模一直是音频信号研究的热点和难点问题[51],然而相关性建模问题至今依然未能得到很好的解决。近年来,以循环神经网络(Recurrent neural networks, RNN)和长短时间记忆(Long short-term memory, LSTM)为代表的深度时序模型的出现,为相关方向的研究带来了新的希望。针对这些开放问题的进一步研究必将推动相关领域的进一步发展。

参考文献:

[1] Virtanen T, Gemmeke J F, Raj B, et al. Compositional models for audio processing:Uncovering the structure of sound mixtures [J]. IEEE Signal Processing Magazine,2015,32(2):125-144.

[2] Gemmeke J F, J. Van De Loo, G. De Pauw, et al. A self-learning assistive vocal interface based on vocabulary learning and grammar induction [C]// 13th Annual Conference of the International Speech Communication Association 2012. Portland, Oregon, USA:International Speech Communication Association (ISCA) Press,2012: 1-4.

[3] Smaragdis P, Brown J C. Non-negative matrix factorization for polyphonic music transcription [C]// IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New Paltz. N Y:IEEE Press, 2003: 177-180.

[4] Guillamet D, Schiele B, Vitria J. Analyzing non-negative matrix factorization for image classification [C]// 16th International Conference on Pattern Recognition. Quebec City, Canada:IEEE Press, 2002:116-119.

[5] Gao Y, Church G. Improving molecular cancer class discovery through sparse non-negative matrix factorization [J]. Bioinformatics, 2005, 21(21):3970-3975.

[6] Févotte C,Dobigeon N. Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization [J]. IEEE Transactions on Image Processing, 2015, 24(12):4810-4819.

[7] 刘维湘, 郑南宁, 游屈波. 非负矩阵分解及其在模式识别中的应用[J]. 科学通报, 2006, 51(3):241-250.
Liu Weixiang, Zheng Nanning, You Qubo. Nonnegative matrix factorization and its applications in pattern recognization [J]. Chinese Science Bulletin, 2006, 51(3):241-250.

[8] 李乐, 章毓晋. 非负矩阵分解算法综述[J]. 电子学报, 2008, 36(4):737-743.
Li Le,Zhang Yujin. A survey on algorithms of non-negative matrix factorization [J]. ACTA Electronica Sinica, 2008, 36(4):737-743.

[9] Lee D D, Seung H S. Learning parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755):788-791.

[10] Basu A, Harris I R, Hjort N L, et al. Robust and efficient estimation by minimising a density power divergence [J]. Biometrika, 1998, 85(3):549-559.

[11] Févotte C, Bertin N, Durrieu J. Nonnegative matrix factorization with the Itakura-Saito divergence:With application to music analysis [J]. Neural Computation, 2009, 21(3):793-830.

[12] Paatero P, Tapper U. Positive matrix factorization:A non-negative factor model with optimal utilization of error estimates of data values [J]. Environmetrics, 1994, 5:111-126.

[13] Lee D D, Seung H S. Algorithms for nonnegative matrix factorization [C]// Advances in Neural Information Processing Systems (NIPS). Denver, CO, USA: MIT Press, 2001:556-562.

[14] Sun D, Févotte C. Alternating direction method of multiplicative for non-negative matrix factorization with beta-divergence [C]// IEEE International Conference on Acoustics, Speech and Signal Processing. Florence, Italy, IEEE Press, 2014:6201-6205.

[15] Guan N, Tao D, Luo Z, et al. NeNMF:An optimal gradient method for nonnegative matrix factorization [J]. IEEE Transactions on Signal Processing, 2012, 60(6):2882-2898.

[16] Cemgil A T. Bayesian inference for nonnegative matrix factorisation models [J]. Computational Intelligence and Neuroscience, 2009,2009:1-17.

[17] Virtanen T. Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(3):1066-1074.

[18] Mohammadiha N, Smaragdis P, Leijon A. Supervised and unsupervised speech enhancement using non-negative matrix fac-

torization [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(10):2140-2151.

[19] 张立伟，贾冲，张雄伟，等. 稀疏卷积非负矩阵分解的语音增强算法[J]. 数据采集与处理，2014，29(2):259-265.
Zhang Liwei, Jia Chong, Zhang Xiongwei, et al. Speech enhancement based on convolutive nonnegative matrix factorization with sparseness constraints[J]. Journal of Data Acquisition and Processing, 2014, 29(2):259-265.

[20] Smaragdis P. Convolutive speech bases and their application to supervised speech separation[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(1):1-12.

[21] Duan Z, Mysore G J, Smaragdis P. Speech enhancement by online non-negative spectrogram decomposition in non-stationary noise environments [C]// 13th Annual Conference of the International Speech Communication Association 2012 (INTERSPEECH 2012). Portland, Oregon, USA: International Speech Communication Association Press, 2012:1-4.

[22] 黄建军，张雄伟，张亚非，等. 时频字典学习的单通道语音增强算法[J]. 声学学报，2012，37(5):539-547.
Huang Jianjun, Zhang Xiongwei, Zhang Yafei, et al. Single channel speech enhancement via time-frequency dictionary learning [J]. ACTA Acustica, 2012, 37(5):539-547.

[23] 李轶南，张雄伟，曾理，等. 改进的稀疏字典学习单通道语音增强算法[J]. 信号处理，2014，30(1):44-50.
Li Yinan, Zhang Xiongwei, Zeng Li, et al. An improved monaural speech enhancement algorithm based on sparse dictionary learning [J]. Signal Processing, 2014, 30(1):44-50.

[24] Sun M, Zhang X, hamme H V, et al. Unseen noise estimation using separable deep auto encoder for speech enhancement [J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2016, 24(1):93-104.

[25] 李轶南，贾冲，杨吉斌，等. 稀疏低秩模型下的单通道自学习语音增强算法[J]. 数据采集与处理，2014，29(2):286-292.
Li Yinan, Jia Chong, Yang Jibin, et al. Self-learning approach for monaural speech enhancement based on sparse and low-rank matrix decomposition [J]. Journal of Data Acquisition and Processing, 2014, 29(2):286-292.

[26] Schmidt M, Larsen J. Reduction of non-stationary noise using a non-negative latent variable decomposition [C]// IEEE Workshop on Machine Learning for Signal Process. (MLSP). Cancun, Mexico:IEEE Press, 2008:486-491.

[27] Li Y, Zhang X, Sun M, et al. Online convolutive non-negative bases learning for speech enhancement [J]. IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences, 2016, E99-A (8):1609-1613.

[28] Sun D L, Mysore G J. Universal speech models for speaker independent single channel source separation [C]// IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Vancouver, British Columbia, Canada:IEEE Press,2013:141-145.

[29] Kim M, Smaragdis P. Mixtures of local dictionaries for unsupervised speech enhancement [J]. IEEE Signal Processing Letters, 2015, 22(3):293-297.

[30] Germain F G, Mysore G J. Speaker and noise independent online single-channel speech enhancement [C]// IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brisbane, Queensland, Australia, IEEE Press,2015:71-75.

[31] Sun M, Li Y, Gemmeke J F, et al. Speech enhancement under low SNR conditions via noise estimation using sparse and low-rank NMF with Kullback-Leibler divergence [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(7):1233-1242.

[32] Li Y, Zhang X, Sun M, et al. Adaptive extraction of repeating non-negative temporal patterns for single-channel speech enhancement [C]// IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Shanghai, China: IEEE Press,2016:494-498.

[33] Li Y, Zhang X, Sun M, et al. Speech enhancement using non-negative low-rank modeling with temporal continuity and sparseness constraints [C]// 17th Pacific-Rim Conference on Multimedia. Xi'an, China: Springer Press, 2016:24-32.

[34] 李轶南，张雄伟，贾冲，等. 稀疏低秩噪声模型下无监督实时单通道语音增强算法[J]. 声学学报，2015，40(4):607-614.
Li Yinan, Zhang Xiongwei, Jia Chong, et al. Unsupervised real-time single channel speech enhancement with low-rank and noise model [J]. Acta Acustica, 2015, 40(4):607-614.

[35] Po-Sen Huang, Scott Deeann Chen, Paris Smaragdis, et al. Singing-voice separation from monaural recordings using robust principal component analysis[C]// IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Kyoto, Japan:[s. n.],2012:57-60.

[36] Wu B, Li K, Yang M, et al. A reverberation-time-aware approach to speech dereverberation based on deep neural networks [J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2017, 25(1):98-107.

[37] Mohammadiha N, Doclo S. Speech dereverberation using non-negative convolutive transfer function and spectro-temporal modeling [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(2):276-289.

[38] Yasuraoka N, Kameoka H, Yoshioka T, et al. I-divergence-based dereveration method with auxiliary function approach

[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. Prague, Czech Republic:IEEE Press, 2011:369-372.

[39] Mohammadiha N, Smaragdis P, Doclo S. Joint acoustic and spectral modeling for speech dereverberation using non-negative representation [C]// IEEE International Conference on Acoustics, Speech and Signal Processing. Brisbane, Queensland, Australia:IEEE Press, 2015:4410-4414.

[40] Singh R, Raj B, Smaragdis P. Latent-variable decomposition based dereverberation of monaural and multi-channel signals [C]// IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). Dallas, Texas, USA:IEEE Press,2010:1914-1917.

[41] Guan N, Tao D, Luo Z,et al. Online non-negative factorization with robust stochastic approximation [J]. IEEE Transactions on Neural Networks and Learning Systems, 2012, 23(7):1087-1099.

[42] Wang D, Vipperla R, Evans N,et al. Online non-negative convolutive pattern learning for speech signals [J]. IEEE Transactions on Signal Processing, 2013:61(1):44-56.

[43] Lefèvre A, Bach F, Févotte C. Online algorithms for nonnegative matrix factorization with the Itakura-Saito divergence [C]// IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. Mohonk Mountain House, New Paltz, N Y, USA:IEEE Press, 2011:313-316.

[44] Yuan M,Lin Y. Model selection and estimation in regression with grouped variables [J]. Journal of the Royal Statistical Society, Series B, 2007, 68(1):49-67.

[45] Zhong M,Girolami M. Reversible jump MCMC for nonnegative matrix factorization [C]// International Conference on Artificial Intelligence and Statistics. Florida, USA:MIT Press,2009:663-670.

[46] Sun M, Zhang X, Hamme H V. A stable approach for model order selection in nonnegative matrix factorization [J]. Pattern Recognition Letters, 2015 (54):97-102.

[47] Tan V Y F, Févotte C. Automatic relevance determination in nonnegative matrix factorization with the beta divergence [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(7):1592-1605.

[48] Li Y, Zhang X, Sun M,et al. Automatic model order selection for convolutive non-negative matrix factorization [J]. IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences, 2016, E99-A (10):1867-1870.

[49] Renkens V, hamme H V. Automatic relevance determination for nonnegative dictionary learning in the Gamma-Poisson model [J]. Signal Processing, 2017, 132:121-133.

[50] Hoffman M, Blei D, Cook P. Bayesian nonparametric matrix factorization for recorded music[C]// 27th International Conference on Machine Learning (ICML 2010). Haifa, Israel: The International Machine Learning Society (IMLS) Press, 2010:439-446.

[51] Smaragdis P, Févotte C, Mysore G J,et al. Static and dynamic source separation using nonnegative matrix factorizations:A unified view[J]. IEEE Signal Processing Magazine, 2014, 31(3):66-75.

作者简介:

张雄伟(1965-),男,教授,博士生导师,研究方向:语音与图像处理、多媒体信息处理,E-mail:xwzhang9898@163.com。

李轶南(1988-),男,博士研究生,研究方向:语音分离、语音增强。

时文华(1982-),女,博士研究生,研究方向:语音与图像处理、多媒体信息处理。

胡永刚(1991-),男,博士研究生,研究方向:多媒体信息处理。

陈栩杉(1987-),男,博士,讲师,研究方向:多媒体信息处理,E-mail:cxs_papip@163.com。