fMRI Time-Series Clustering Using a Mixture of Mixtures of Student's-*t* and Rayleigh Distributions

Qunyi Xie, Zhuyan Zhang, Xu Pan, Hongqing Zhu*
School of Information Science & Engineering, East China University of Science and Technology, Shanghai 200237, China
*Email: hqzhu@ecust.edu.cn

Abstract—**In this paper, a new Markov random field-based mixture model, where each of its components is a mixture of Student's-*t* and Rayleigh distributions, is proposed for clustering fMRI time-series. By introducing the non-symmetric Rayleigh distribution, the proposed algorithm has flexibility to fit various types of observed time-series. Moreover, our method incorporates Markov random field so that the spatial relationships between neighboring voxels are considered, which makes the presented model more robust to noise, and that preserves more details of the clustering results compared with other symmetric distribution-based algorithms. Additionally, the expectation maximization algorithm is directly implemented to estimate the parameter set by maximizing the data log-likelihood function. The proposed framework is evaluated on real fMRI time-series, and the quantitatively compared results are demonstrated in terms of effectiveness and accuracy.**

I. Introduction

Functional magnetic resonance imaging (fMRI) provides a non-invasive technique in both neuro-imaging and clinical fields [1-2]. During the past decades, several approaches have been proposed for detecting activation cortex with fMRI time-series of the brain [3]. Generally, the fMRI time-series with similar temporal behavior belong to the same class, in contrast, dissimilar ones should belong to different classes. Clustering technique is usually performed to partition unlabeled time-series into various classes. Thus, seeking for a robust and accurate clustering method is becoming the key to analyze fMRI time-series.

Methods proposed for fMRI time-series analysis mainly include model-based and data driven approaches. In model-based approaches, Gaussian mixture model (GMM) provides a simple and effective way for time-series clustering [4]. Nevertheless, in consideration of its sensitivity to outliers, recently, an advanced Student's-*t* mixture mode (SMM) [5-6] has been employed. Unlike GMM, SMM with longer tails has the degrees of freedom which provides a more flexible method than GMM. However, the main drawback of the two mixture models is that they can't take the spatial information into account. Therefore, both GMM and SMM are extremely sensitive to noise. To overcome this shortcoming, mixture models based on Markov random field (MRF) have been widely used recently, but this could be time expensive. In addition, each component of these MRF-based mixture models is a symmetric distribution [7-8]. In real fMRI time-series, the intensity distribution of each class does not show exact symmetric. This leads to undesirable clustering results while a symmetric distribution is used. More recently, Browne et al. [9] introduced a model-based learning method using a mixture of Gaussian and uniform distributions. However, this algorithm also adopts a symmetric probability distribution function. Based on these considerations, this paper proposes a novel MRF-based mixture model of mixtures of Student's-*t* and Rayleigh distribution (RSMM-MRF). The main advantage of the proposed algorithm is that it has capability to model the non-symmetric intensity distribution due to the introduction of the non-symmetric Rayleigh distribution. In addition, the proposed model utilizes MRF for each pixel to impose spatial smoothness constraints. Furthermore, in our method, a new smoothing prior is introduced in MRF distribution to reduce the computational complexity. Finally, this study adopts the expectation maximization (EM) algorithm for maximizing the data log-likelihood function, and obtaining the estimations of all parameters. Numerical simulations demonstrate that the proposed method presents more accurate clustering results for fMRI time-series than other related methods.

The rest of this paper is organized as follows. In section II, we briefly describe the standard Student's-*t* mixture model. In section III, the details of our proposed method are presented, followed by the process of parameter learning described in section IV. Experiments conducted on the real fMRI time-series are demonstrated in section V. Section VI gives the conclusions.

II. Standard Student's-T Mixture Model

Let $x_i, i = (1, 2, ..., N)$, denote an observation at the ith pixel of an image, and $(\Omega_1, \Omega_2, ..., \Omega_K)$ denote different classes. Student's-*t* mixture model assumes that the density function $p(x_i|\Theta)$ at each pixel x_i is given by

$$p(x_i|\Theta) = \sum_{j=1}^{K} \pi_j f(x_i|\mu_j, \Sigma_j, v_j), \quad (1)$$

where the prior probability π_j satisfies the following constrains

$$0 \leq \pi_j \leq 1 \quad \text{and} \quad \sum_{j=1}^{K} \pi_j = 1, \quad (2)$$

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

and $\Theta = \{\pi_j, \mu_j, \Sigma_j, v_j\}$ is the parameter set of the Student's-t mixture model. Each Student's-t distribution $f(x_i|\mu_j, \Sigma_j, v_j)$ in (1) is written in the following form

$$f(x_i|\mu_j, \Sigma_j, v_j) = \frac{\Gamma(\frac{v_j+D}{2})|\Sigma_j|^{-\frac{1}{2}}}{(\pi v_j)^{\frac{D}{2}}\Gamma(\frac{v_j}{2})[1+v_j^{-1}\delta(x_i, \mu_j; \Sigma_j)]^{\frac{v_j+D}{2}}}, \tag{3}$$

where D is the dimensionality of the observation x_i, and μ_j,Σ_j,v_j are mean, covariance, and the degrees of freedom, respectively. $\delta(x_i, \mu_j; \Sigma_j)$ is a simple expression for $(x_i - \mu_j)^T\Sigma_j^{-1}(x_i - \mu_j)$. The log-likelihood function of the density function (1) can be derived as

$$\begin{aligned} L(\Theta) &= \log\left\{\prod_{i=1}^{N} p(x_i|\Theta)\right\} \\ &= \sum_{i=1}^{N}\log\left\{\sum_{j=1}^{K}\pi_j f\left(x_i|\mu_j, \Sigma_j, v_j\right)\right\}. \end{aligned} \tag{4}$$

III. Proposed Method

Let $Y = \{y_1, y_2, ..., y_N\}$ be a set of N fMRI time-series, where the T-dimensional observation y_i represents a sequence of T values measured over time, i.e., $y_i = \{y_{it}, t = 1, ..., T\}$. To model fMRI time-series, we use the linear regression model to assume y_i having the following form

$$y_i = X\beta_i + e_i, \tag{5}$$

where β_i is the P-dimensional unknown vector of linear regression coefficients, and the last term e_i denotes unknown stochastic noise. This paper chooses the first P-order DCT orthogonal polynomials with length T as the design matrix $X = \{x_{tk}, t = 1, ..., T, k = 1, ..., P\}$, which is denoted by

$$x_{tk} = \cos\left(\pi(k-1)\frac{2t-1}{2T}\right). \tag{6}$$

In this study, we consider partitioning the fMRI time-series Y into K classes. The observations belonging to the same class satisfy the same conditional probability distribution. Thus, each class has its own parameter set θ_j, and the proposed finite mixture model can be written as

$$p\left(y_i|\Pi, \Theta\right) = \sum_{j=1}^{K}\pi_{ij} f\left(y_i|\theta_j\right), \tag{7}$$

where $\Pi = \{\pi_{ij}, i = 1, ..., N, j = 1, ..., K\}$ is the set of prior probability, and $\Theta = [\theta_1, \theta_2, ..., \theta_K]$ is the parameter set of the mixture model. The proposed component density function $f(y_i|\theta_j)$ in (7) applies a mixture of Student's-t and Rayleigh distributions, which has the following form

$$f\left(y_i|\theta_j\right) = w_j\phi\left(y_i|\beta_j, \Sigma_j, v_j\right) + (1-w_j)\, R\left(y_i|\lambda_j\right), \tag{8}$$

where $w_j \in [0, 1]$ is a weight factor, and $\phi(y_i|\beta_j, \Sigma_j, v_j)$ is the density function of T-dimensional multivariate Student's-t distribution with the following form

$$\phi(y_i|\beta_j, \Sigma_j, v_j) = \frac{\Gamma(\frac{v_j+T}{2})|\Sigma_j|^{-\frac{1}{2}}}{(\pi v_j)^{\frac{T}{2}}\Gamma(\frac{v_j}{2})[1+v_j{}^{-1}\delta(y_i, \beta_j; \Sigma_j)]^{\frac{v_j+T}{2}}}, \tag{9}$$

where Σ_j is diagonal covariance matrix and $\delta(y_i, \beta_j; \Sigma_j)$ is defined by $(y_i - X\beta_j)^T\Sigma_j^{-1}(y_i - X\beta_j)$. According to [10], a multivariate Rayleigh distribution $R(y_i|\lambda_j)$ is defined as

$$R\left(y_i|\lambda_j\right) = \frac{||y_i||}{\lambda_j}\exp\left(-\frac{{y_i}^T y_i}{2\lambda_j}\right), \tag{10}$$

where $\lambda_j \in (0, \infty)$ is the scale parameter and $||\cdot||$ denotes L_1 norm.

In order to take the spatial information into account, MRF is introduced to the proposed algorithm. Based on Bayes' rules, the posterior probability density function of our method can be expressed as

$$p\left(\Pi, \Theta|Y\right) \propto p\left(Y|\Pi, \Theta\right) p\left(\Pi\right). \tag{11}$$

Combing (7) and (8), the joint density function of the time-series Y in (11) can be written as

$$\begin{aligned} p\left(Y|\Pi, \Theta\right) &= \prod_{i=1}^{N} p\left(y_i|\Pi, \Theta\right) \\ &= \prod_{i=1}^{N}\sum_{j=1}^{K}\pi_{ij}\left[w_j\phi\left(y_i|\beta_j, \Sigma_j, v_j\right) + (1-w_j)\, R\left(y_i|\lambda_j\right)\right]. \end{aligned} \tag{12}$$

Here, the Markov random field based on the Gibbs distribution is denoted by

$$p\left(\Pi\right) = Z^{-1}\exp\left\{-\frac{1}{\bar{T}}U\left(\Pi\right)\right\}, \tag{13}$$

where Z and $\bar{T}$ are normalizing constants. To simplify the proposed model and make it computationally efficient, a new smoothing prior $U(\Pi)$ is applied in current study as follows:

$$U\left(\Pi\right) = -\sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)}\log\pi_{ij}^{(t+1)}, \tag{14}$$

where t indicates the iteration step and the function $G_{ij}^{(t)}$ in (14) is defined as

$$G_{ij}^{(t)} = \exp\left[\frac{\alpha}{2N_i}\sum_{m\in\bar{N}_i}\left(z_{mj}^{(t)} + \pi_{mj}^{(t)}\right)\right], \tag{15}$$

where $z_{mj}^{(t)}$ and $\pi_{mj}^{(t)}$ respectively denote the posterior and prior probability. α is a smoothness controlling parameter ($\alpha = 12$). N_i is the number of time-series in the neighborhood $\bar{N}_i$ of the observation y_i. Hence, the MRF distribution $p(\Pi)$ in (13) can be rewritten as

$$p\left(\Pi\right) = Z^{-1}\exp\left\{\frac{1}{\bar{T}}\sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)}\log\pi_{ij}^{(t+1)}\right\}. \tag{16}$$

Considering (12) and (16), the log-likelihood function of (11) can be denoted by

$$\begin{aligned} L(\Pi,\Theta|Y) &= \log\{p(\Pi,\Theta|Y)\} \\ &= \sum_{i=1}^{N}\log\{\sum_{j=1}^{K}\pi_{ij}^{(t+1)}[w_j^{(t+1)}\phi(y_i|\beta_j^{(t+1)},\Sigma_j^{(t+1)},v_j^{(t+1)}) \\ &+ (1-w_j^{(t+1)})R(y_i|\lambda_j^{(t+1)})]\} - \log Z \\ &+ \frac{1}{\overline{T}}\sum_{i=1}^{N}\sum_{j=1}^{K}G_{ij}^{(t)}\log\pi_{ij}^{(t+1)}. \end{aligned} \tag{17}$$

IV. Parameter Learning

The objective of this section is to optimize the parameter set. Note that, we can not directly apply the EM algorithm for maximizing the log-likelihood function (17). To overcome this problem, Jensen's inequality is introduced so that two hidden variables z_{ij} and η_{ij} are computed in E-Step. Given the values of normalizing constants Z=1 and $\overline{T}$=1, the objective function can be obtained by

$$\begin{aligned} J(\Pi,\Theta|Y) &= \sum_{i=1}^{N}\sum_{j=1}^{K}z_{ij}^{(t)}\{\log\pi_{ij}^{(t+1)} \\ &+ \eta_{ij}^{(t)}\log w_j^{(t+1)} + \eta_{ij}^{(t)}\log\phi(y_i|\beta_j^{(t+1)},\Sigma_j^{(t+1)},v_j^{(t+1)}) \\ &+ (1-\eta_{ij}^{(t)})\log(1-w_j^{(t+1)}) + (1-\eta_{ij}^{(t)})\log R(y_i|\lambda_j^{(t+1)})\} \\ &+ \sum_{i=1}^{N}\sum_{j=1}^{K}G_{ij}^{(t)}\log\pi_{ij}^{(t+1)}, \end{aligned} \tag{18}$$

where the posterior probability z_{ij} can be computed as follows

$$z_{ij}^{(t)} = \frac{\pi_{ij}^{(t)}f\left(y_i|\theta_j^{(t)}\right)}{\sum_{m=1}^{K}\pi_{im}^{(t)}f\left(y_i|\theta_m^{(t)}\right)}, \tag{19}$$

and another parameter η_{ij} is formulated as

$$\eta_{ij}^{(t)} = \frac{w_j^{(t)}\phi\left(y_i|\beta_j^{(t)},\Sigma_j^{(t)},v_j^{(t)}\right)}{w_j^{(t)}\phi\left(y_i|\beta_j^{(t)},\Sigma_j^{(t)},v_j^{(t)}\right)+\left(1-w_j^{(t)}\right)R\left(y_i|\lambda_j^{(t)}\right)}. \tag{20}$$

To estimate the prior probability, we calculate the partial derivative of the objective function with respect to π_{ij}

$$\frac{\partial}{\partial\pi_{ij}^{(t+1)}}\left[J-\sum_{i=1}^{N}\gamma_i\left(\sum_{j=1}^{K}\pi_{ij}^{(t+1)}-1\right)\right]=0, \tag{21}$$

where γ_i is Lagrange's multiplier. Since π_{ij} satisfies the constraints $0\le\pi_{ij}\le 1$ and $\sum_{j=1}^{K}\pi_{ij}=1$, the prior probability is calculated by

$$\pi_{ij}^{(t+1)} = \frac{z_{ij}^{(t)}+G_{ij}^{(t)}}{\sum_{m=1}^{K}\left(z_{im}^{(t)}+G_{im}^{(t)}\right)}. \tag{22}$$

Similarly, taking the derivative of objective function $\partial J/\partial w_j$ as zero, we have

$$w_j^{(t+1)} = \frac{\sum_{i=1}^{N}z_{ij}^{(t)}\eta_{ij}^{(t)}}{\sum_{i=1}^{N}z_{ij}^{(t)}}. \tag{23}$$

Setting the partial derivative of the objective functions with respect to β_j to zero, yields

$$\begin{aligned} \beta_j^{(t+1)} &= \left[\sum_{i=1}^{N}z_{ij}^{(t)}\eta_{ij}^{(t)}u_{ij}^{(t)}X^T\Sigma_j^{-1(t)}X\right]^{-1} \\ &\times X^T\Sigma_j^{-1(t)}\sum_{i=1}^{N}z_{ij}^{(t)}\eta_{ij}^{(t)}u_{ij}^{(t)}y_i, \end{aligned} \tag{24}$$

where $u_{ij}^{(t)}$ is defined as

$$u_{ij}^{(t)} = \frac{v_j^{(t)}+T}{v_j^{(t)}+\left(y_i-X\beta_j^{(t)}\right)^T\Sigma_j^{-1(t)}\left(y_i-X\beta_j^{(t)}\right)}. \tag{25}$$

To obtain the estimation of variance Σ_j, we compute the $\partial J/\partial\Sigma_j=0$, and obtain the following expression

$$\Sigma_{jl}^{(t+1)} = \frac{\sum_{i=1}^{N}z_{ij}^{(t)}\eta_{ij}^{(t)}u_{ij}^{(t)}\left(y_{il}-\left[X\beta_j^{(t+1)}\right]_l\right)^2}{\sum_{i=1}^{N}z_{ij}^{(t)}\eta_{ij}^{(t)}}. \tag{26}$$

Note that $l=1,2,...,T$ denotes the lth diagonal element. The value of v_j can be calculated as a solution to the following equation

$$\begin{aligned} &\frac{\sum_{i=1}^{N}z_{ij}^{(t)}\eta_{ij}^{(t)}\left(\log u_{ij}^{(t)}-u_{ij}^{(t)}\right)}{\sum_{i=1}^{N}z_{ij}^{(t)}\eta_{ij}^{(t)}}+\log\left(\frac{v_j^{(t+1)}}{2}\right)-\varphi\left(\frac{v_j^{(t+1)}}{2}\right) \\ &+\varphi\left(\frac{v_j^{(t)}+T}{2}\right)-\log\left(\frac{v_j^{(t)}+T}{2}\right)+1=0, \end{aligned} \tag{27}$$

where $\varphi(x)=\partial(\ln\Gamma(x))/\partial x$ is the digamma function. With respect to the scaled parameter λ_j of Rayleigh distribution, the necessary condition for making its partial derivative of the objective function zero becomes

$$\lambda_j^{(t+1)} = \frac{\sum_{i=1}^{N}z_{ij}^{(t)}\left(1-\eta_{ij}^{(t)}\right)y_i^Ty_i}{2\sum_{i=1}^{N}z_{ij}^{(t)}\left(1-\eta_{ij}^{(t)}\right)}. \tag{28}$$

Thus, the discussion on estimating parameter set can be summarized as follows.

Step 1. Initialize the prior probability π_{ij}, the weighting factor w_j, the regression coefficients β_j, the covariance matrix Σ_j, the degrees of freedom v_j, and the scaled parameter λ_j. Let α=12.

Step 2. Compute the posterior probability z_{ij}, the parameter

η_{ij}, the new function G_{ij}, and the variable u_{ij} with (19), (20), (15), and (25), respectively.
Step 3. Update the prior probability π_{ij}, the weighting factor w_j, the regression coefficients β_j, the covariance matrix Σ_j, the degrees of freedom v_j, and the scaled parameter λ_j using (22), (23), (24), (26), (27) and (28) respectively.
Step 4. Terminate the iterations if the objective function converges. Otherwise, $t = t + 1$ and return to Step 2.

V. Experimental Results

In this section, the algorithms are verified on two real fMRI time-series, including auditory and attention data. Since no ground truth is available, the clustering results are visually compared with different algorithms. In this experiment, real fMRI datasets obtained by the SPM12 package [11] are preprocessed by the following standard steps, i.e., realignment, segmentation, normalization, spatial smoothing, and scaling with the mean value of the data. All these details are described in the SPM manual.

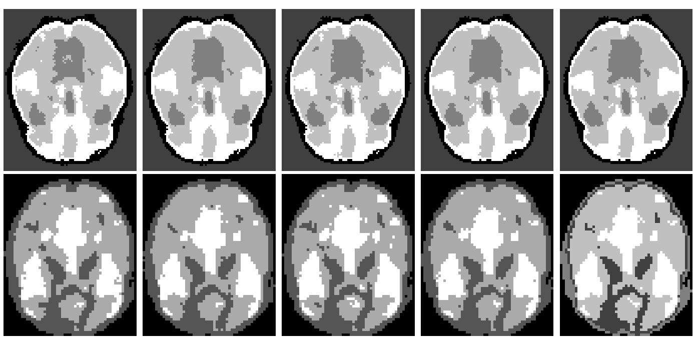

Fig. 1. Clustering results (K=5) using auditory data (First Row) and attention data (Second Row) with Gaussian Noise (Mean = 0, Variance = 0.01). From left to right are GMM, GMM-MRF, SMM, SMM-MRF, and RSMM-MRF methods.

The experiment first evaluates the proposed scheme using the real auditory and attention datasets. The former is collected from a healthy volunteer which consists of 96 acquisitions with $7s$ duration. After pre-processing, the functional images used in the current experiment are composed of 79 slices ($79 \times 59 \times 79$ voxels). Here, we select slice 29 for this experiment. The latter with resolution $53 \times 63 \times 46$ is obtained from echo-planar imaging on a 2 Tesla Magnetom Vision MRI system. Slice 18 is used to compare the effection of RSMM-MRF. In addition, we have applied GMM, GMM-MRF [7], SMM and SMM-MRF [8] to analyze these datasets. Fig.1 shows the clustering results of activation areas. As shown in the final clustering results, with respect to the methods with MRF, one can found that the effect of noise tends to disappear as well as small islands of activation area. Furthermore, the proposed RSMM-MRF prevents the clustering regions from losing details under noisy conditions, and obtains more attractive results.
In order to quantitatively assess the clustering results, the intra-label error is adopted in this experiment. The intra-label error is defined by

$$\varepsilon_l = \frac{1}{N} \sum_{j=1}^{K} \sum_{i=1}^{N_j} ||y_i - X\beta_j||^2, \tag{29}$$

where N_j denotes the number of time-series in class Ω_j. The lower ε_l is, the more accurate clustering result the algorithm has. The intra-label errors of various models are then calculated under different noisy cases. The experiment compares several methods with the proposed model on the real fMRI time-series listed in Tables I and II. As can be seen, the proposed method can recognize the details from the noise using the mixtures of Student's-t and Rayleigh distributions. Another interesting discovery of this experiment is that, the clustering results of GMM-MRF and SMM-MRF with the attention dataset perform poorer than GMM and SMM. It is may because that the small islands of activation area in attention dataset are regarded as noise and smoothed by MRF.

TABLE I
Intra-label error of different methods on an auditory processing task

Gaussian Noise Mean	Variance	GMM	GMM-MRF	SMM	SMM-MRF	RSMM-MRF
0	0.005	0.5216	0.5213	0.4756	0.4764	0.4676
0	0.008	0.5338	0.5327	0.4936	0.4931	0.4852
0	0.01	0.5397	0.5397	0.5029	0.5028	0.4931
0	0.02	0.5836	0.5798	0.5571	0.5517	0.5418
0.01	0.005	0.6338	0.6326	0.5792	0.5773	0.5701
0.01	0.008	0.6442	0.6401	0.6024	0.5971	0.5891
0.01	0.01	0.6502	0.6540	0.6094	0.6087	0.6022
0.01	0.02	0.6958	0.6933	0.6704	0.6681	0.6643

TABLE II
Intra-label error of different methods on an attention processing task

Gaussian Noise Mean	Variance	GMM	GMM-MRF	SMM	SMM-MRF	RSMM-MRF
0	0.005	0.1878	0.1865	0.1868	0.1860	0.1856
0	0.008	0.1928	0.1920	0.1890	0.1894	0.1879
0	0.01	0.1973	0.1979	0.1933	0.1934	0.1924
0	0.02	0.2258	0.2266	0.2164	0.2193	0.2154
0.01	0.005	0.2136	0.2138	0.2067	0.2075	0.2058
0.01	0.008	0.2204	0.2212	0.2136	0.2135	0.2122
0.01	0.01	0.2243	0.2249	0.2167	0.2198	0.2172
0.01	0.02	0.2564	0.2578	0.2461	0.2467	0.2459

In statistics, paired t-test [12-13] is applied to further evaluate the significant difference between the proposed method and the others. In this hypothesis testing method, an original hypothesis that the two algorithms have no significant difference is pre-established firstly and then tested by a statistic satisfying the student's-t distribution. The statistic can be computed by the data analysis tool. In our experiment, 10 trials with noisy fMRI datasets via various methods are performed to obtain their intra-label errors. The results of paired t-test based on the intra-label errors are shown in Table III. As shown in this table,

TABLE III
fMRI DATASET-CLUSTERING ACCURACY FOR THE 10 TRIALS WITH NOISY AUDITORY AND ATTENTION DATASETS IN TERMS OF INTRA-LABEL ERROR (MEAN ± STANDARD). BOLD = LOWEST VALUE,

Clustering Methods	Auditory	Attention
GMM	$0.60034 \pm 0.24585^{**}$	$0.21480 \pm 0.14456^{**}$
GMM-MRF	$0.59919 \pm 0.24548^{**}$	$0.21509 \pm 0.14702^{**}$
SMM	$0.56133 \pm 0.25066^{**}$	$0.20858 \pm 0.13497^{*}$
SMM-MRF	$0.55940 \pm 0.24856^{**}$	$0.20945 \pm 0.13704^{**}$
RSMM-MRF	$\mathbf{0.55168} \pm 0.25147$	$\mathbf{0.20780} \pm 0.13613$

*= Values significantly different from RSMM-MRF (paired t-test with 0.05 significance level);
**= Values significantly different from RSMM-MRF (paired t-test with 0.01 significance level).

the results indicate that the proposed method is significantly different from the classic methods. Therefore, the proposed improving of the clustering method is notable.

VI. CONCLUSIONS

In this paper, a novel mixture model using mixtures of Student's-t and Rayleigh distributions for clustering fMRI time-series has been proposed. By taking the spatial relationships into account, the proposed method demonstrates better robustness against noise. The EM algorithm is employed to achieve parameters learning. The clustering results of real fMRI datasets including auditory and attention data confirm robustness and accuracy of the proposed model and its significant difference with the classic methods.

ACKNOWLEDGMENT

The authors would like to thank the National Nature Science Foundation of China under Grant 61371150.

REFERENCES

[1] C. Bchel and K. J. Friston, "Modulation of connectivity in visual pathways by Attention: cortical interactions evaluated with structural equation modelling and fMRI," *Cerebral Cortex*, vol.7, no.8, pp.768-778, 1997.

[2] V. P. Oikonomou and K. Blekas, "An adaptive regression mixture model for fMRI cluster analysis," *IEEE. Trans. Medical Imaging*, vol.32, no.4, pp.649-659, 2013.

[3] V. P. Oikonomou, B. Konstantinos, and A. Loukas, "A sparse and spatially constrained generative regression model for fMRI data analysis," *IEEE. Biomed. Eng and Trans*, vol.59, no.1, pp.58-67, 2012.

[4] K. Blekas, N. Galatsanos, and A. Likas, "A Sparse regression mixture model for clustering time-series," *Artificial Intelligence: Theories, Models and Applications*, Springer Berlin Heidelberg, pp.64-72, 2008.

[5] G. Sfikas, C. Nikou, and N. Galatsanos, "Robust image segmentation with mixtures of Student's t-distributions," in *Proc. IEEE Int. Conf. Image Process*. IEEE, 2007, vol. 1, pp. 273-276.

[6] M. N. Thanh, Q. M. J. Wu, and H. Zhang, "A nonsymmetric mixture model for unsupervised image segmentation," *IEEE Trans. Cybern*, vol. 43, no.2, pp. 751-765, 2013.

[7] M. N. Thanh and Q. M. J. Wu, "Gaussian-mixture-model-based spatial neighborhood relationships for pixel labeling problem," *IEEE. Trans. Systems Man and Cybernetics Society*, vol.42, no.1, pp.193-202, 2012.

[8] H. Zhang, Q. M. J. Wu, and T. M. Nguyen, "Modified student's t-hidden Markov model for pattern recognition and classification," *IET. Signal. Processing*, vol. 7, pp. 219-227, 2013.

[9] R. P. Browne, P. D. Mcnicholas, and M. D. Sparling, "Model-based learning using a mixture of mixtures of Gaussian and uniform distributions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.34, no.4, pp.814-817, 2012.

[10] J. Seabra, F. Ciompi, and O. Pujol, "Rayleigh mixture model for plaque characterization in intravascular ultrasound," *IEEE. Trans. Biomed. Eng*, vol.58, no.5, pp.1314-1324, 2011.

[11] K. Friston. (2009). *Statistical parameter mapping*. [online]: Available: http://www.fil.ion.ucl.ac.uk/spm/.

[12] A. Ribbens, J. Hermans, and F. Maes, "Unsupervised segmentation, clustering, and groupwise registration of heterogeneous populations of brain MR images," *IEEE. Trans. Medical Imaging*, vol.33, no.2, pp.201-224, 2014.

[13] S. Jain, D. M. Sima, and A. Ribbens, "Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images," *Neuroimage Clinical*, vol.8, pp.367-375, 2015.