IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON AI-DRIVEN BIG DATA PROCESSING: THEORY, METHODOLOGY, AND APPLICATIONS

Received February 28, 2019, accepted March 10, 2019, date of publication March 21, 2019, date of current version April 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2906669

Positive Data Modeling Using a Mixture of Mixtures of Inverted Beta Distributions

YUPING LAI[1], (Member, IEEE), XIU MA[1], (Student Member, IEEE), YANPING XU[2], (Member, IEEE), YONGFA LING[3], CHUNLAI DU[1], JIANHE DU[4], YONGMEI ZHANG[1], AND YUAN PING[5]

[1]School of Computer Science and Technology, North China University of Technology, Beijing 100144, China
[2]School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 311305, China
[3]School of Information and Communication Engineering, Hezhou University, Hezhou 542899, China
[4]School of Information and Communication Engineering, Communication University of China, Beijing 100024, China
[5]School of Information Engineering, Xuchang University, Xuchang 461000, China

Corresponding authors: Yanping Xu (xuyanping@hdu.edu.cn) and Yongmei Zhang (zhangym@ncut.edu.cn)

This work was supported in part by the Beijing Natural Science Foundation under Grant 4194076 and Grant 4172019, in part by the Beijing Municipal Education Commission Science and Technology Plan under Grant KM201910009014, in part by the Joint of Beijing Natural Science Foundation and Education Commission under Grant KZ201810009011, in part by the Program for Science and Technology Innovation Talents in Universities of Henan Province under Grant 18HASTIT022, and in part by the National Natural Science Foundation of China under Grant 61802094 and Grant 61601414.

ABSTRACT Finite mixture models based on the symmetric Gaussian distribution have been applied broadly in data analysis. However, not all the data in real-world applications can be safely supposed to have a symmetric Gaussian form. This paper presents a new mixture model that includes the inverted Beta mixture model (IBMM) as a special case to analyze the positive non-Gaussian data. The advantage of the proposed model is that the number of the model parameters is variable and infinite. Consequently, the proposed model is adaptable to the size of the data. On the basis of the recently proposed extended variational inference (EVI) framework, we develop a closed-form solution to approximate the posterior distributions. The performance and the effectiveness of the proposed model are demonstrated with the real data generated from two challenging applications, namely, image classification and object detection.

INDEX TERMS Bayesian estimation, extended variational inference, inverted Beta distribution, image classification, object detection.

I. INTRODUCTION

Finite mixture models [1]–[9] provide a powerful and flexible statistical tool for analyzing heterogeneous data arising from multiple populations. These models have been widely applied to study a large number of important problems in various domains, such as machine learning, data mining and computer vision. The finite Gaussian mixture model (GMM) [1], [2], [10], [11] has been the most widely applied approach in real-world applications. Modeling the statistics of observations via the GMM has two important advantages. First, the GMM can model an arbitrary continuous distribution with an appropriate number of components. Second, the parameters in a GMM can be efficiently estimated via maximum likelihood (ML) estimation with the expectation-maximization (EM) algorithm. The Gaussian distribution is symmetric and unbounded (with support range $(-\infty, +\infty)$); however, the observed data in many practical applications are not symmetric or unbounded.

Recently, non-Gaussian statistical models, such as the Beta mixture model (BMM) [5], the Dirichlet mixture model (DMM) [3], [6], [7], the Beta-Liouville mixture model (BLMM) [12], and the Watson mixture model (WMM) [8], have gained considerable attention since they can provide better modeling capabilities than the GMM in the case of non-Gaussian data [3], [7], [12], [14]–[16], such as bounded or semibounded support data. For example, in the area of signal processing, the power spectrum, which is the most frequently used feature, is semibounded in the range $(0, +\infty)$ and asymmetric. In the area of computer vision, image normalized histograms and bag-of-words representations of images are bounded with support range $[0, 1]$. Also, in the area of speech transmission, the line spectral frequency (LSF) representation of the linear predictive coding parameters is bounded in the limited range $[0, \pi]$.

To overcome this problem, the IBMM has been proposed in [13]. Compared to the Gaussian distribution, the inverted

The associate editor coordinating the review of this manuscript and approving it for publication was Zhanyu Ma.

Beta distribution has a flexible shape and can be symmetric or highly skewed. However, there are two major disadvantages of the IBMM. First, the EM algorithm can result in overfitting when it is excessively complex. Second, the iterative numerical estimation is prohibitively expensive and timeconsuming for practical applications. To address these problems, a Bayesian estimation method for the IBMM based on the multiple lower bound (MLB) approximation has been proposed in [13]. Because the MLB approximation cannot guarantee convergence, the Bayesian estimation method for IBMM using the single lower bound (SLB) approximation has been proposed in [6] and [17] to handle this problem. For more details with respect to (w.r.t.) the SLB and MLB approximation, interested readers are referred to [6], [17], and [18]. Compared to the EM algorithm for the IBMM, the Bayesian estimation method has several advantages. First, the overfitting problem can be prevented. Furthermore, it can simultaneously select the optimal mixture of components to estimate the model parameters. Finally, an analytically tractable solution for the true posterior distribution can be obtained in this inference process; therefore, it may scale well to large applications. One drawback of all the aforementioned mixture models is that their distributions are unbounded with support range.

Another way to fit different shapes of the non-Gaussian data, which models each mixture component with multiple distributions, has attracted considerable attention. Examples of such models previously studied in the literature include the mixture of mixtures of Gaussian distributions [19], [20], mixture of mixtures of Student's t-distributions [21], and mixtures of Gaussian and uniform distributions [22]. One drawback of these mixture models is that their distributions are unbounded with support range $(-\infty, +\infty)$. However, not all the data we would like to model can be safely supposed to be unbounded [3], [5]–[8], [12], [13]. To overcome this problem, the mixture of mixtures of Beta distributions has been proposed in [21] to model bounded data. However, this method is used only for single-dimensional data. Moreover, the Bayesian estimation method with MLB approximation has been used to learn the model parameters, which cannot guarantee the convergence of the proposed algorithm [6], [17], [18].

Motivated by the aforementioned observations, in this paper, we propose a novel finite mixture model to model the probability density function (PDF) of positive data. Notably, our approach differs from those discussed above. First, a mixture of mixtures of Beta distributions is used to model only the univariate bounded data; we overcome this problem by proposing a new mixture model that has multivariate inverted Beta components for D-dimensional positive data. Second, the SLB approximation-based method is adopted to optimize the parameters. The proposed method is used in two important applications, namely, image classification and object detection, and the good performance is verified via real data evaluations.

The rest of this paper is organized as follows. In Section 2, we present the proposed method in detail. In Section 3, a Bayesian learning algorithm with EVI is derived. The experimental results on real data are reported in Section 4. Finally, some conclusions are drawn in Section 5.

II. PROPOSED METHOD

A D-dimensional random vector $\mathbf{x} = [\mathrm{x}_1, \cdots, \mathrm{x}_D]^{\mathrm{T}}$ is said to have a parametric finite mixture distribution if its PDF is written in the form

$$p(\mathbf{x}|\boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{m=1}^{M} \pi_m p(\mathbf{x}|\boldsymbol{\vartheta}_m), \tag{1}$$

where $\boldsymbol{\pi} = [\pi_1, \cdots, \pi_M]^{\mathrm{T}}$ denotes the mixture weights, which satisfy the following constraints:

$$0 \leq \pi_m \leq 1, \quad \text{and} \sum_{m=1}^{M} \pi_m = 1, \tag{2}$$

$\boldsymbol{\theta} = \{\boldsymbol{\vartheta}_1, \cdots, \boldsymbol{\vartheta}_M\}$ is the parameter vector, and $p(\mathbf{x}|\boldsymbol{\vartheta}_m)$ is called the "component density". For more comprehensive reviews w.r.t. probabilistic mixing modeling, interested readers are referred to [24]. How to choose the component density $p(\mathbf{x}|\boldsymbol{\vartheta}_m)$ is at the heart of mixture modeling. Note that $p(\mathbf{x}|\boldsymbol{\vartheta}_m)$ can be any type of distribution. The majority of mixture models select the Gaussian distribution, Student's t-distribution, or generalized Gaussian distribution as the component density. These three distributions are symmetric and unbounded. However, the observed data from many real-world applications are not symmetric or unbounded. To address this problem, a new parametric finite mixture distribution is presented, where each component density is modeled by multiple multivariate inverted Beta distributions [13]

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\alpha}_m, \boldsymbol{\beta}_m, \boldsymbol{\eta}_m) &= \sum_{k=1}^{K_m} \eta_{mk} \mathbf{iBeta}(\mathbf{x}|\boldsymbol{\alpha}_{mk}, \boldsymbol{\beta}_{mk}) \\ &= \sum_{k=1}^{K_m} \eta_{mk} \prod_{d=1}^{D} \mathrm{iBeta}(\mathrm{x}_d|\alpha_{mkd}, \beta_{mkd}), \end{aligned} \tag{3}$$

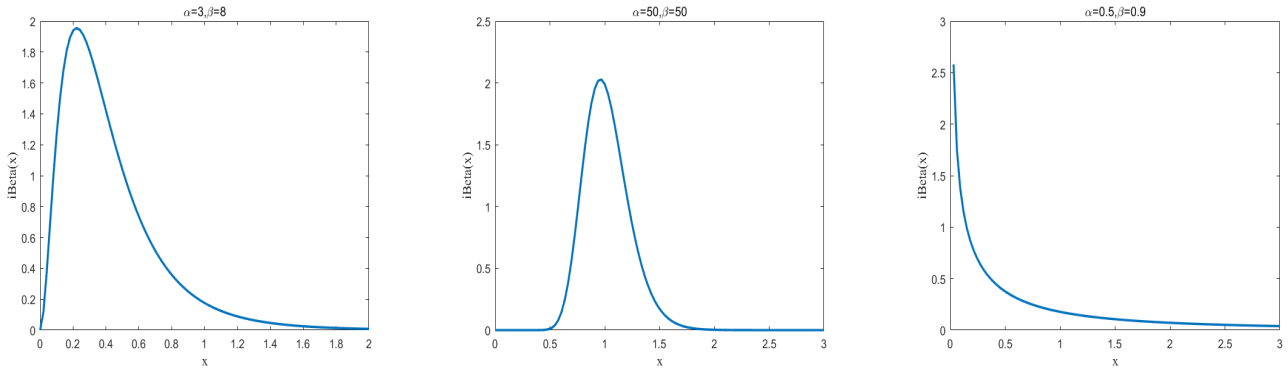
where $\boldsymbol{\eta}_m = [\eta_{m1}, \cdots, \eta_{mK_m}]^{\mathrm{T}}$ is called the mixing weighting factors, which satisfy the following conditions:

$$0 \leq \eta_{mk} \leq 1, \quad \text{and} \sum_{k=1}^{K_m} \eta_{mk} = 1, \tag{4}$$

$\boldsymbol{\alpha}_m = \{\alpha_{mdk}\}$ and $\boldsymbol{\beta}_m = \{\beta_{mdk}\}$ are the parameter sets of the mth mixture component and K_m is the number of the multivariate inverted Beta distributions used to model the mth mixture component in (3). In addition, $\mathrm{iBeta}(\mathrm{x}, \alpha, \beta)$ denotes the inverted Beta distribution

$$\mathrm{iBeta}(\mathrm{x}\ \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \mathrm{x}^{\alpha-1}(1+\mathrm{x})^{-(\alpha+\beta)}, \quad \alpha,\ \beta > 0, \tag{5}$$

where $\mathrm{x} > 0$ and $\Gamma(\cdot)$ is the Gamma function. The shape of the inverted Beta distribution relies on two shape parameters,



(a) (b) (c)

FIGURE 1. Inverted Beta distributions for different pairs of parameters.

α and β, which could be symmetric or highly skewed. Some typical cases are shown in Fig. 1.

Assume that $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N]$ is a set of observations that are drawn independently from the mixture distribution given by (1). The likelihood is then written as

$$p(\mathbf{X}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\eta}) = \prod_{n=1}^{N} \left\{ \sum_{m=1}^{M} \pi_m \left[\sum_{k=1}^{K_m} \eta_{mk} \prod_{d=1}^{D} \text{iBeta}(\text{x}_{nd}|\alpha_{mkd}, \beta_{mkd}) \right] \right\}. \tag{6}$$

The ML framework with EM can be utilized to estimate the model parameters by maximizing the usual quantity in (6) over the parameters. However, this objective function cannot be utilized to perform our model size selection. Thus, the Bayesian estimation method based on variational inference (VI) [4] is adopted to address this problem.

For convenience, the proposed model is formulated as a latent variable model. For each observation $\mathbf{x}_n$, a latent indicator vector $\mathbf{z}_n = \{\text{z}_{n1}, \cdots, \text{z}_{nM}\}$ is introduced to represent component membership, with $\text{z}_{nm} \in \{0, 1\}$ and such that $\text{z}_{nm} = 1$ if $\mathbf{x}_n$ is generated from component m and $\text{z}_{nm} = 0$ otherwise. The weighting factor η_{nm} is expressed by the latent variables s_{nmk}, with $\text{s}_{nmk} \in \{0, 1\}$, such that $s_{nmk} = 1$ if $\mathbf{x}_n$ is drawn from the mth component in (1) and $\mathbf{iBeta}(\mathbf{x}_n|\boldsymbol{\alpha}_{mk}, \boldsymbol{\beta}_{mk})$ is associated with the kth component in (3) and $s_{nmk} = 0$ otherwise. The latent variable model of the proposed model is then specified as

$$p(\mathbf{Z}|\boldsymbol{\pi}) = \prod_{n=1}^{N} \prod_{m=1}^{M} \pi_m^{\text{z}_{nm}}, \tag{7}$$

$$p(\mathbf{S}|\boldsymbol{\eta}) = \prod_{n=1}^{N} \prod_{m=1}^{M} \prod_{k=1}^{K_m} \eta_{mk}^{\text{s}_{nmk}}, \tag{8}$$

$$p(\mathbf{X}, \mathbf{Z}, \mathbf{S}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{n=1}^{N} \prod_{m=1}^{M} \left\{ \prod_{k=1}^{K_m} \left[\prod_{d=1}^{D} \text{iBeta}(\text{x}_{nd}|\alpha_{mdk}, \beta_{mdk}) \right]^{\text{s}_{nmk}} \right\}^{\text{z}_{nm}}. \tag{9}$$

To obtain the Bayesian formulation of the proposed model, suitable prior distributions need to be placed over the model parameters. Since the inverted Beta distribution is a member of the exponential family, it has a formal conjugate prior. However, the prior is not analytically tractable and cannot be directly applied for VI [6]. Herein, the conjugate priors for u_{mkd} and v_{mkd} are selected to be Gamma priors

$$p(\boldsymbol{\alpha}|\mathbf{u}, \mathbf{v}) = \prod_{m=1}^{M} \prod_{k=1}^{K_m} \prod_{d=1}^{D} \mathcal{G}(\alpha_{mkd}|u_{mkd}, v_{mkd}), \tag{10}$$

$$p(\boldsymbol{\beta}|\mathbf{g}, \mathbf{h}) = \prod_{m=1}^{M} \prod_{k=1}^{K_m} \prod_{d=1}^{D} \mathcal{G}(\beta_{mkd}|g_{mkd}, h_{mkd}). \tag{11}$$

where $\mathcal{G}(\cdot)$ stands for a Gamma distribution. Note that $\boldsymbol{\pi}$ and $\boldsymbol{\eta}$ are viewed as parameters rather than stochastic variables, so no priors are placed over them.

For notational simplicity, we denote $\boldsymbol{\Theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{Z}, \mathbf{S}\}$ as the set of random variables and $\boldsymbol{\Lambda} = \{\boldsymbol{\pi}, \boldsymbol{\eta}\}$ as the set of parameters. The joint probability of the model is then given by

$$\begin{aligned} p(\mathbf{X}, \boldsymbol{\Theta}|\boldsymbol{\Lambda}) &= p(\mathbf{X}, \mathbf{Z}, \mathbf{S}|\boldsymbol{\alpha}, \boldsymbol{\beta}) p(\mathbf{Z}|\boldsymbol{\pi}) p(\mathbf{S}|\boldsymbol{\eta}) p(\boldsymbol{\alpha}) p(\boldsymbol{\beta}) \\ &= \prod_{n=1}^{N} \prod_{m=1}^{M} \left\{ \prod_{k=1}^{K_m} \left[\prod_{d=1}^{D} \frac{\Gamma(\alpha_{mkd} + \beta_{mkd})}{\Gamma(\alpha_{mkd})\Gamma(\beta_{mkd})} \text{x}_{nd}^{\alpha_{mkd}-1} \right. \right. \\ &\quad \left. \left. \times (1 + \text{x}_{nd})^{-(\alpha_{mkd}+\beta_{mkd})} \right]^{\text{s}_{nmk}} \right\}^{\text{z}_{nm}} \\ &\quad \times \left[\prod_{n=1}^{N} \prod_{m=1}^{M} \pi_m^{\text{z}_{nm}} \right] \left[\prod_{n=1}^{N} \prod_{m=1}^{M} \prod_{k=1}^{K_m} \eta_{mk}^{\text{s}_{nmk}} \right] \\ &\quad \times \prod_{m=1}^{M} \prod_{k=1}^{K_m} \prod_{d=1}^{D} \frac{v_{mkd}^{u_{mkd}}}{\Gamma(u_{mkd})} \alpha_{mkd}^{u_{mkd}-1} e^{-v_{mkd}\alpha_{mkd}} \\ &\quad \times \prod_{m=1}^{M} \prod_{k=1}^{K_m} \prod_{d=1}^{D} \frac{h_{mkd}^{g_{mkd}}}{\Gamma(g_{mkd})} \beta_{mkd}^{g_{mkd}-1} e^{-h_{mkd}\beta_{mkd}}. \end{aligned} \tag{12}$$

A graphical illustration of the proposed model is presented in Fig. 2. In the following section, we estimate the model parameters $\boldsymbol{\Theta}$ and select the optimal number of components $\{M, K_m\}$.

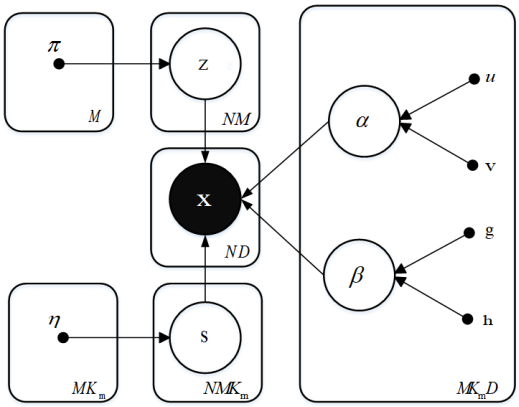


FIGURE 2. Graphical representation of the Bayesian model. Symbols in circles denote stochastic variables; all others are hyperparameters. Arrows show the relationships between variables. Rounded boxes denote the repetitions (with the number of repetitions in the corners).

III. PARAMETER LEARNING

In this section, an EVI-based Bayesian estimation approach is presented to simultaneously solve the problems of parameter estimation and model size selection for our proposed model. The proposed variational framework is able to overcome the overfitting problem.

A. VARIATIONAL BAYESIAN INFERENCE

Following the approach proposed in [25], the aforementioned two issues can be simultaneously solved by maximizing the ML $p(\mathbf{X}|\mathbf{\Lambda})$ w.r.t. $\mathbf{\Theta}$

$$p(\mathbf{X}|\mathbf{\Lambda}) = \int p(\mathbf{X}, \mathbf{\Theta}|\mathbf{\Lambda}) d\mathbf{\Theta}. \tag{13}$$

Notably, the integral in (13) denotes joint integration over $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{Z}\}$ and summation over $\{\mathbf{S}, \mathbf{\Pi}\}$. Due to the coupling among these variables, this marginalization is intractable. Thus, a variational method is used to find a tractable lower bound over $p(\mathbf{X}|\mathbf{\Lambda})$ by introducing an arbitrary variational distribution $q(\mathbf{\Theta})$ to approximate the true posterior distribution $p(\mathbf{\Theta}|\mathbf{X}, \mathbf{\Lambda})$ and using a popular decomposition for $\ln p(\mathbf{X}|\mathbf{\Lambda})$

$$\ln p(\mathbf{X}|\Lambda) = \mathrm{KL}(q||p) + \mathcal{L}(q), \tag{14}$$

where $\mathrm{KL}(q||p) = \int q(\mathbf{\Theta}) \ln\{q(\mathbf{\Theta})/p(\mathbf{\Theta}|\mathbf{X}, \mathbf{\Lambda})\} d\mathbf{\Theta}$ is the Kullback-Leibler (KL) divergence between $q(\mathbf{\Theta})$ and $p(\mathbf{\Theta}|\mathbf{X}, \mathbf{\Lambda})$ and $\mathcal{L}(q) = \int q(\mathbf{\Theta}) \ln\{p(\mathbf{\Theta}, \mathbf{X}|\mathbf{\Lambda})/q(\mathbf{\Theta})\} d\mathbf{\Theta}$. Since $\mathrm{KL}(q||p)$ is nonnegative and would become exact if $q(\mathbf{\Theta}) = p(\mathbf{\Theta}|\mathbf{X}, \mathbf{\Lambda})$, $\mathcal{L}(q)$ is a lower bound of $\ln p(\mathbf{X}|\mathbf{\Lambda})$. Maximizing the lower bound $\mathcal{L}(q)$ is the same as minimizing the KL divergence $\mathrm{KL}(q||p)$. However, minimizing $\mathrm{KL}(q||p)$ to solve $q(\mathbf{\Theta})$ is infeasible since $p(\mathbf{\Theta}|\mathbf{X}, \mathbf{\Lambda})$ is unknown. Hence, optimization of $\mathcal{L}(q)$ has been extensively employed in the VI framework to achieve a good variational approximation. The lower bound $\mathcal{L}(q)$ is known as the variational objective function and can be rewritten as

$$\mathcal{L}(q) = \mathrm{E}_q[\ln p(\mathbf{X}, \mathbf{\Theta}|\mathbf{\Lambda})] - \mathrm{E}_q[q(\mathbf{\Theta})]. \tag{15}$$

Nevertheless, $\mathcal{L}(q)$ is not available in closed form since the expectation $\mathrm{E}_q[\ln p(\mathbf{X}, \mathbf{\Theta}|\mathbf{\Lambda})]$ is not tractable. The recently proposed EVI framework is adopted to address this problem. The main idea behind this framework is that if we can further bound $\mathcal{L}(q)$ as

$$\mathcal{L}(q) \geq \tilde{\mathcal{L}}(q) = \mathrm{E}_q[\ln \tilde{p}(\mathbf{X}, \mathbf{\Theta}|\mathbf{\Lambda})] - \mathrm{E}_q[q(\mathbf{\Theta})], \tag{16}$$

where $\tilde{p}(\mathbf{X}, \mathbf{\Theta}|\mathbf{\Lambda})$ is a "help function" that is subject to the constraint $\mathrm{E}_q[\ln p(\mathbf{X}, \mathbf{\Theta}|\mathbf{\Lambda})] \geq \mathrm{E}_q[\ln \tilde{p}(\mathbf{X}, \mathbf{\Theta}|\mathbf{\Lambda})]$, then we can asymptotically approach the optimal solution by maximizing $\tilde{\mathcal{L}}(q)$. Moreover, to yield a tractable expression for $\tilde{\mathcal{L}}(q)$, we have to restrict the form of $q(\mathbf{\Theta})$. It is convenient to assume a variational distribution $q(\mathbf{\Theta})$ with a factorized form as

$$q(\mathbf{\Theta}) = \left[\prod_{n=1}^{N}\prod_{m=1}^{M} q(\mathrm{z}_{nm})\right]\left[\prod_{n=1}^{N}\prod_{m=1}^{M}\prod_{k=1}^{K_m} q(\mathrm{s}_{nmk})\right] \times \left[\prod_{m=1}^{M}\prod_{k=1}^{K_m}\prod_{d=1}^{D} q(\alpha_{mkd}) q(\beta_{mkd})\right]. \tag{17}$$

Through simple calculation, it is straightforward to obtain the approximate optimal solution as

$$\ln q_s(\Theta_s) = \langle \ln \tilde{p}(\mathbf{X}, \mathbf{\Theta}|\mathbf{\Lambda}) \rangle_{j \neq s} + \text{Const}, \tag{18}$$

where $\langle \cdot \rangle_{j \neq s}$ denotes the expectation w.r.t. all factors $q_j(\Theta_j)$ except for $j = s$ and "Const" denotes a constant that is applied to normalize the corresponding distribution [4]. Because the expression of factor $q_s(\Theta_s)$ has some functional dependence on the expectations computed w.r.t. the other factors $q_j(\Theta_j)$ for $j \neq s$, these factors do not represent an explicit solution. This limitation requires cycling of each factor to find the maximum value of $\tilde{\mathcal{L}}(q)$. To conduct this optimization, every factor $q_s(\Theta_s)$ must be suitably initiated in advance and then updated in turn by a revised estimate given by (21). Since $\tilde{\mathcal{L}}(q)$ is convex w.r.t. each factor, convergence is theoretically guaranteed for $q_s(\Theta_s)$ [4], [17], [18].

B. VARIATIONAL POSTERIOR DISTRIBUTIONS

This section details how we can apply (18) to compute the variational posterior distributions in (17). First, we need to compute the expectation of $\ln p(\mathbf{X}, \mathbf{\Theta}|\mathbf{\Lambda})$ as

$$\begin{aligned}
&\langle \ln p(\mathbf{X}, \mathbf{\Theta}|\mathbf{\Lambda}) \rangle \\
&= \sum_{n=1}^{N}\sum_{m=1}^{M} \langle \mathrm{z}_{nm} \rangle \left\{ \sum_{k=1}^{K_m} \langle \mathrm{s}_{nmk} \rangle \sum_{d=1}^{D} [\mathcal{R}_{mkd} \right. \\
&\quad \left. + (\langle \alpha_{mkd} \rangle - 1) \ln \mathrm{x}_{nd} - (\langle \alpha_{mkd} \rangle + \langle \beta_{mkd} \rangle) \ln(1 + \mathrm{x}_{nd})] \right\} \\
&\quad + \sum_{n=1}^{N}\sum_{m=1}^{M} \langle \mathrm{z}_{nm} \rangle \ln \pi_m + \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{k=1}^{K_m} \langle \mathrm{s}_{nmk} \rangle \ln \eta_{nmk} \\
&\quad + \sum_{m=1}^{M}\sum_{k=1}^{K_m}\sum_{d=1}^{D} [(u_{mkd} - 1)\langle \ln \alpha_{mkd} \rangle - v_{mkd} \langle \alpha_{mkd} \rangle] \\
&\quad + \sum_{m=1}^{M}\sum_{k=1}^{K_m}\sum_{d=1}^{D} [(g_{mkd} - 1)\langle \ln \beta_{mkd} \rangle - h_{mkd} \langle \beta_{mkd} \rangle] \\
&\quad + \text{Const},
\end{aligned} \tag{19}$$

where we define

$$\mathcal{R}_{mkd} = \left\langle \ln \frac{\Gamma(\alpha_{mkd} + \beta_{mkd})}{\Gamma(\alpha_{mkd})\Gamma(\beta_{mkd})} \right\rangle. \tag{20}$$

Since evaluation of $\langle \ln p(\mathbf{X}, \mathbf{\Theta}|\mathbf{\Lambda}) \rangle$ requires calculation of $\mathcal{R}_{mkd}$, which is not available in closed form, optimization of each variational factor in (17) is infeasible via the formal VI framework. The abovementioned EVI framework is adopted to address this problem. According to [6, eq. (25)], $\mathcal{R}_{mkd}$ can be bounded as

$$\begin{aligned}\mathcal{R}_{mkd} &\geq \tilde{\mathcal{R}}_{mkd} \\ &= \ln \frac{\Gamma(\bar{\alpha}_{mkd} + \bar{\beta}_{mkd})}{\Gamma(\bar{\alpha}_{mkd})\Gamma(\bar{\beta}_{mkd})} + [\Psi(\bar{\alpha}_{mkd} + \bar{\beta}_{mkd}) \\ &\quad - \Psi(\bar{\alpha}_{mkd})](\langle \ln \alpha_{mkd} \rangle - \ln \bar{\alpha}_{mkd})\bar{\alpha}_{mkd} \\ &\quad + [\Psi(\bar{\alpha}_{mkd} + \bar{\beta}_{mkd}) - \Psi(\bar{\beta}_{mkd})] \\ &\quad \times (\langle \ln \beta_{mkd} \rangle - \ln \bar{\beta}_{mkd})\bar{\beta}_{mkd}, \end{aligned} \tag{21}$$

where $\Psi(\cdot)$ represents the digamma function. Substituting (21) back into (19), $\langle \ln p(\mathbf{X}, \mathbf{\Theta}|\mathbf{\Lambda}) \rangle$ is bounded as

$$\begin{aligned}&\langle \ln \tilde{p}(\mathbf{X}, \mathbf{\Theta}|\mathbf{\Lambda}) \rangle \\ &= \sum_{n=1}^{N}\sum_{m=1}^{M} \langle \mathrm{z}_{nm} \rangle \left\{ \sum_{k=1}^{K_m} \langle \mathrm{s}_{nmk} \rangle \sum_{d=1}^{D} \Big[\tilde{\mathcal{R}}_{mkd} + (\langle \alpha_{mkd} \rangle - 1) \ln \mathrm{x}_{nd} \right. \\ &\quad \left. - (\langle \alpha_{mkd} \rangle + \langle \beta_{mkd} \rangle) \ln(1 + \mathrm{x}_{nd}) \Big] \right\} \\ &\quad + \sum_{n=1}^{N}\sum_{m=1}^{M} \langle \mathrm{z}_{nm} \rangle \ln \pi_m + \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{k=1}^{K_m} \langle \mathrm{s}_{nmk} \rangle \ln \eta_{nmk} \\ &\quad + \sum_{m=1}^{M}\sum_{k=1}^{K_m}\sum_{d=1}^{D} [(u_{mkd} - 1)\langle \ln \alpha_{mkd} \rangle - v_{mkd} \langle \alpha_{mkd} \rangle] \\ &\quad + \sum_{m=1}^{M}\sum_{k=1}^{K_m}\sum_{d=1}^{D} [(g_{mkd} - 1)\langle \ln \beta_{mkd} \rangle - h_{mkd} \langle \beta_{mkd} \rangle] \\ &\quad + \text{Const}. \end{aligned} \tag{22}$$

By combining (18) and (22), it is straightforward to derive the optimal variational solutions for $q(\mathbf{Z})$, $q(\mathbf{S})$, $q(\boldsymbol{\alpha})$ and $q(\boldsymbol{\beta})$, as follows.

1) OPTIMAL SOLUTION TO $q(\mathbf{Z})$

Absorbing all terms that are independent of z_{nm} into a constant yields

$$\begin{aligned}&\ln q^*(\mathrm{z}_{nm}) \\ &= \mathrm{z}_{nm} \left\{ \ln \pi_m + \sum_{k=1}^{K_m} \langle \mathrm{s}_{nmk} \rangle \sum_{d=1}^{D} [\tilde{\mathcal{R}}_{mkd} + (\langle \alpha_{mkd} \rangle - 1) \right. \\ &\quad \left. \times \ln \mathrm{x}_{nd} - (\langle \alpha_{mkd} \rangle + \langle \beta_{mkd} \rangle) \ln(1 + \mathrm{x}_{nd})] \right\} + \text{Const}. \end{aligned} \tag{23}$$

Taking the exponential of both sides of (23), we have

$$\ln q^*(\mathbf{Z}) = \sum_{n=1}^{N}\sum_{m=1}^{M} \mathrm{z}_{nm} \ln \rho_{nm} + \text{Const}, \tag{24}$$

Algorithm 1 Bayesian Estimation of the Proposed Model

1: Set the number of components M and K_m.
2: Initialize the parameters: u_{mkd}, v_{mkd}, g_{mkd}, h_{mkd}.
3: Initialize r_{nm} and λ_{nmk} by K-means algorithm.
4: **repeat**
5: **E step:** Update responsibilities r_{nm}, λ_{nmk} by (26) and (29).
6: **M step:** Update variational parameters u^*_{mkd}, v^*_{mkd}, g^*_{mkd}, h^*_{mkd} by (33), (34), (36), (37).
7: **until** Stop criterion is reached.
8: Determine M and K_m by discarding the components that have small mixture weights ($\leq 10^{-5}$).

where ρ_{nm} is given by

$$\begin{aligned}\ln \rho_{nm} &= \ln \pi_m + \sum_{k=1}^{K_m} \langle \mathrm{s}_{nmk} \rangle \sum_{d=1}^{D} [\tilde{\mathcal{R}}_{mkd} + (\langle \alpha_{mkd} \rangle - 1) \\ &\quad \times \ln \mathrm{x}_{nd} - (\langle \alpha_{mkd} \rangle + \langle \beta_{mkd} \rangle) \ln(1 + \mathrm{x}_{nd})]. \end{aligned} \tag{25}$$

Hence, we obtain

$$q^*(\mathbf{Z}) = \prod_{n=1}^{N}\prod_{m=1}^{M} r_{nm}^{\mathrm{z}_{nm}}, \; r_{nm} = \frac{\rho_{nm}}{\sum_{m=1}^{M} \rho_{nm}}, \tag{26}$$

where $q^*(\mathbf{Z})$ is a categorical distribution. For $q^*(\mathbf{Z})$, we have $\langle \mathrm{z}_{nm} \rangle = r_{nm}$.

2) OPTIMAL SOLUTION TO $q(\mathbf{S})$

Considering terms involving only s_{nmk}, we obtain

$$\begin{aligned}&\ln q^*(\mathrm{s}_{nmk}) \\ &= \mathrm{s}_{nmk} \left\{ \ln \eta_{mk} + \sum_{d=1}^{D} [\tilde{\mathcal{R}}_{mkd} + (\langle \alpha_{mkd} \rangle - 1) \ln \mathrm{x}_{nd} \right. \\ &\quad \left. - (\langle \alpha_{mkd} \rangle + \langle \beta_{mkd} \rangle) \ln(1 + \mathrm{x}_{nd}) \right] \Big\} + \text{Const}. \end{aligned} \tag{27}$$

Taking the exponential of both sides of (27), we have

$$q^*(\mathbf{S}) = \prod_{n=1}^{N}\prod_{m=1}^{M}\prod_{k=1}^{K_m} \lambda_{nmk}^{s_{nmk}}, \tag{28}$$

where λ_{nmk} is given by

$$\lambda_{nmk} = \frac{\delta_{nmk}}{\sum_{k=1}^{K_m} \delta_{nmk}}, \tag{29}$$

where we define

$$\begin{aligned}\ln \delta_{nmk} &= \ln \eta_{mk} + \langle \mathrm{z}_{nm} \rangle \sum_{d=1}^{D} [\tilde{\mathcal{R}}_{mkd} + (\langle \alpha_{mkd} \rangle - 1) \\ &\quad \times \ln \mathrm{x}_{nd} - (\langle \alpha_{mkd} \rangle + \langle \beta_{mkd} \rangle) \ln(1 + \mathrm{x}_{nd})]. \end{aligned} \tag{30}$$

For $q^*(\mathbf{S})$, we have $\langle \mathrm{s}_{nmk} \rangle = \lambda_{nmk}$.

FIGURE 3. Example images from the Caltech-05 dataset. (a) Face. (b) Motorbike. (c) Airplane. (d) Car side. (e) Background.

FIGURE 4. Sample images from the Scene-13 dataset. (a) Office. (b) Living room. (c) Bedroom. (d) Kitchen. (e) Suburb. (f) Tall building. (g) Street. (h) Inside of city.

3) OPTIMAL SOLUTION TO $q(\alpha)$

Absorbing the terms that do not have some functional dependence on α_{mkd} into a constant, we have

$$\begin{aligned}\ln q^*(\alpha_{mkd}) = \ln \alpha_{mkd} \Bigg\{ \sum_{n=1}^{N} \langle z_{nm} \rangle \langle s_{nmk} \rangle \, \bar{\alpha}_{mkd} \\ \times [\Psi(\bar{\alpha}_{mkd} + \bar{\beta}_{mkd}) - \Psi(\bar{\alpha}_{mkd})] + u_{mkd} - 1 \Bigg\} \\ + \alpha_{mkd} \left[v_{mkd} - \sum_{n=1}^{N} \langle z_{nm} \rangle \langle s_{nmk} \rangle \ln x_{nd} \right] + \text{Const.}\end{aligned} \tag{31}$$

Taking the exponential of both sides of (31), $q^*(\boldsymbol{\alpha})$ is recognized to be a Gamma density

$$q^*(\boldsymbol{\alpha}) = \prod_{m=1}^{M} \prod_{k=1}^{K_m} \prod_{d=1}^{D} \mathcal{G}(\alpha_{mkd} | u^*_{mkd}, v^*_{mkd}). \tag{32}$$

The hyperparameters u^*_{mkd} and v^*_{mkd} in (32) are given by

$$u^*_{mkd} = \sum_{n=1}^{N} \langle z_{nm} \rangle \langle s_{nmk} \rangle [\Psi(\bar{\alpha}_{mkd} + \bar{\beta}_{mkd}) - (\bar{\alpha}_{mkd})] \bar{\alpha}_{mkd}, \tag{33}$$

$$v^*_{mkd} = v_{mkd} - \sum_{n=1}^{N} \langle z_{nm} \rangle \langle s_{nmk} \rangle [\ln x_{nd} - \ln (1 + x_{nd})]. \tag{34}$$

4) OPTIMAL SOLUTION TO $q(\beta)$

The posterior distribution $q^*(\boldsymbol{\beta})$ is recognized to be a Gamma density

$$q^*(\boldsymbol{\beta}) = \prod_{m=1}^{M} \prod_{k=1}^{K_m} \prod_{d=1}^{D} \mathcal{G}(\beta_{mkd} | g^*_{mkd}, h^*_{mkd}), \tag{35}$$

where g^*_{mkd} and h^*_{mkd} are the hyperparameters

$$g^*_{mkd} = \sum_{n=1}^{N} \langle z_{nm} \rangle \langle s_{nmk} \rangle [\Psi(\bar{\alpha}_{mkd} + \bar{\beta}_{mkd}) - \Psi(\bar{\beta}_{mkd})] \bar{\beta}_{mkd}, \tag{36}$$

TABLE 1. Comparison of the image classification accuracies (measured in %) obtained by different methods. The p-values of the Student's t-test under the null hypothesis, i.e., that our method and the compared method have equal means but unknown variances, are listed.

	Our method	IBMM [17]	IDMM [26]	SMM [27]
Caltech-05	**94.57**(0.68)	92.21(0.57)	89.14(1.08)	86.81(0.63)
p-values	N/A	2.53×10^{-4}	8.61×10^{-4}	4.36×10^{-7}
Scene-13	**91.73**(1.44)	89.69(1.47)	86.13(1.18)	84.82(1.09)
p-values	N/A	7.89×10^{-3}	4.12×10^{-3}	1.92×10^{-5}

$$h^*_{mkd} = h_{mkd} + \sum_{n=1}^{N} \langle z_{nm} \rangle \langle s_{nmk} \rangle \ln(1 + x_{nd}). \tag{37}$$

The other expectations involved in (21) are given by

$$\begin{aligned} \bar{\alpha}_{mkd} &= u^*_{mkd}/v^*_{mkd}, \quad \langle \ln \alpha_{mkd} \rangle = \Psi(u^*_{mkd}) - \ln v^*_{mkd}, \\ \bar{\beta}_{mkd} &= g^*_{mkd}/h^*_{mkd}, \quad \langle \ln \beta_{mkd} \rangle = \Psi(g^*_{mkd}) - \ln h^*_{mkd}. \end{aligned} \tag{38}$$

By setting the derivatives of $\tilde{\mathcal{L}}(q)$ w.r.t. π_m and η_{mk} to zero, we have

$$\pi_m = \frac{1}{N} \sum_{n=1}^{N} r_{nm}, \eta_{mk} = \frac{1}{N} \sum_{n=1}^{N} \lambda_{nmk}. \tag{39}$$

Alg. 1 presents a summary of the algorithm. Since the lower bound is convex w.r.t. each factor, the optimization procedure is guaranteed to converge [4]. The effectiveness of the proposed approach is demonstrated in the next section.

IV. EXPERIMENTAL RESULTS

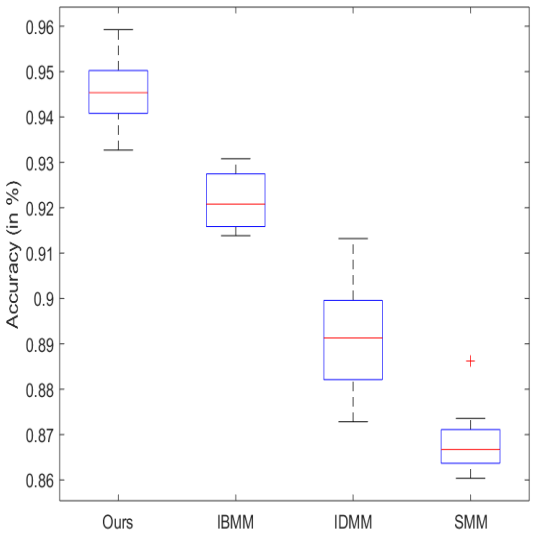
A. DESIGN OF EXPERIMENTS

The main purpose of our experiments is to compare the performance of the proposed mixture model to that of the IBMM [17], IDMM [26], and SMM [27] in terms of positive data modeling. To ensure a fair comparison, the model parameters are estimated by a VI-based approach. For details w.r.t. IDMM and IBMM, the reader is referred to [13] and [27], respectively. In this section, we focus on applications involving positive data, namely, image classification and object detection. Note that comparison of our results with those of the latest approaches proposed for these two tasks is out of the scope of this paper. The actual purpose is to evaluate mixture-based methods. In these applications, the number of components M and K_m are initialized to 15 and 5, respectively. To offer non-informative priors, we set the prior distribution parameters as $u_{mdk} = g_{mdk} = 1$ and $v_{mdk} = h_{mdk} = 0.1$. The posterior means are taken as the point estimates of the parameters.

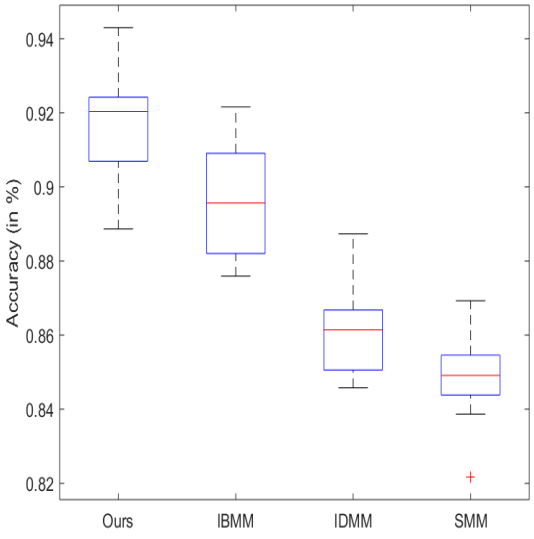
B. IMAGE CLASSIFICATION

Image classification is a very important research topic in computer vision that has attracted substantial attention. The aim of image classification is to classify an image into a predefined category [29]–[32]. Image classification has a variety of applications, such as traffic scene recognition [33]–[35], medical image mining [36]–[38], and intelligent video analysis [39]–[42]. Although it is easy for humans to perform this task, it remains challenging for computers owing to pose, illustration and scale variations, occlusions and intra-class variability. In recent years, a large number of research efforts have been devoted to overcoming such difficulties. These efforts have focused mainly on the study of excellent image descriptors and robust and efficient classifiers. In this section, the rectangular histogram of oriented gradient (R-HOG) descriptor [43] is used to represent each image, and our proposed model is used as a classifier.

The evaluations were conducted on two datasets. The first dataset is the Caltech-05 dataset, which contains the



(a)



(b)

FIGURE 5. Boxplots for comparison of the distributions of the classification accuracies. The central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles. (a) Caltech-05. (b) Indoor/outdoor.

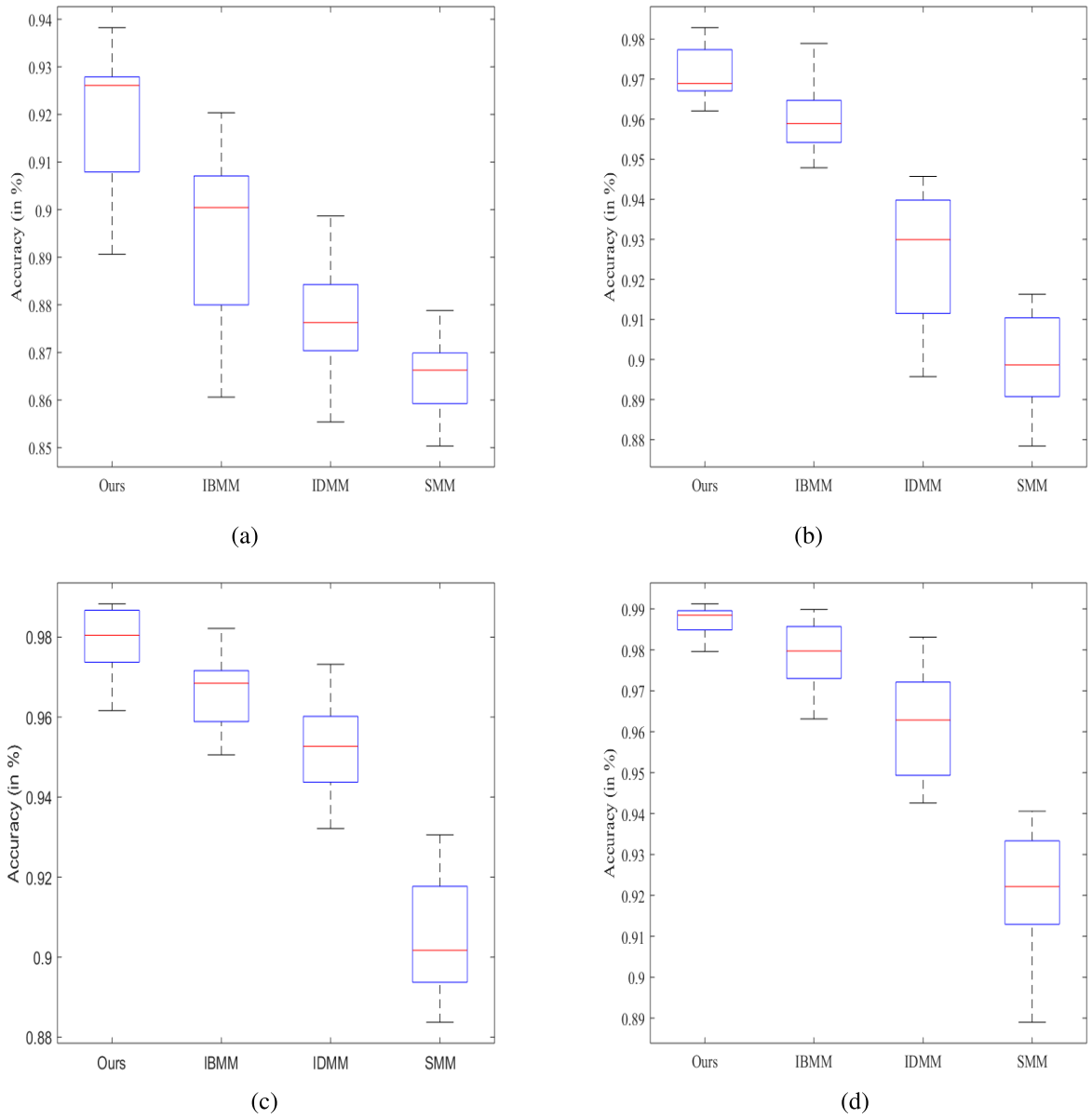


FIGURE 6. Boxplots for comparisons of the distributions of the detection accuracies. The central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles. (a) Face. (b) Motorbike. (c) Airplane. (d) Car side.

five most popular categories from the Caltech-101 dataset[1]: faces (435 images), motorbikes (798 images), airplanes (800 images), car sides (123 images), and backgrounds (467 images). Fig. 3 shows example images from these five categories. The second dataset is the indoor/outdoor dataset, which contains two categories: indoor scenes and outdoor scenes. The indoor scenes are composed of four subcategories from the Scene-13 dataset[2]: office (216 images), living room (289 images), bedroom (174 images), and kitchen (151 images). The outdoor scenes consist of four subcategories from the Scene-13 dataset: suburb residence (241 images), tall buildings (356 images), streets (292 images), and inside of cities (308 images). Example images from each category are shown in Fig. 4. During the evaluations, each category was randomly divided 15 times into two equal parts, one for training and the other for testing.

The proposed method for image classification can be summarized as follows. First, experiments were conducted by setting the number of windows and histogram bins to 7 and 9, respectively, such that each image was represented by a 441-dimensional positive feature vector. Second, the feature vectors of each category in the training sets were modeled by the proposed model. Finally, Bayes' rule was employed to allocate the testing image to a given category according to the posterior probabilities. Tab. 1 shows the average classification accuracies along with the standard deviations. Fig. 5 illustrates the distributions of the

[1]http://www.vision.caltech.edu/archive.html
[2]http://vision.stanford.edu/Datasets/AnimTransDistr.rar

TABLE 2. Comparison of the image classification accuracies (measured in %) obtained by different methods. The p-values of the Student's t-test under the null hypothesis, i.e., that our method and the compared method have equal means but unknown variances, are listed.

	Our method	IBMM [17]	IDMM [26]	SMM [27]
Faces	**91.84**(1.49)	89.41(1.84)	87.75(1.05)	86.48(0.83)
p-values	N/A	5.64×10^{-3}	4.39×10^{-3}	2.01×10^{-6}
Motorbikes	**97.12**(0.59)	95.98(0.79)	92.65(1.60)	89.95(1.09)
p-values	N/A	0.043	7.14×10^{-4}	1.08×10^{-6}
Airplanes	**97.98**(0.74)	96.67(0.93)	95.19(1.05)	90.51(1.36)
p-values	N/A	5.32×10^{-3}	2.41×10^{-5}	6.94×10^{-5}
Car sides	**98.69**(0.35)	97.85(0.76)	96.08(1.23)	92.02(1.52)
p-values	N/A	0.068	1.46×10^{-3}	2.33×10^{-6}

classification accuracies. These results indicate that the proposed method outperforms the other three methods.

C. OBJECT DETECTION

Object detection is an extremely challenging problem with a massive amount of important applications, such as intelligent monitoring [44], [45], intelligent traffic systems [46], [47], and face image research [48]–[50]. The main purpose of this task is to distinguish a specific object from all possible objects. Over the past few years, object detection has received considerable research focus in the area of computer vision [51], [52]. Like most computer vision tasks, a critical step for accurate object detection is to extract good descriptors to represent object images. Here, the R-HOG descriptor with seven windows and nine histogram bins was employed. Thus, each image in the dataset was represented by a 441-dimensional positive feature vector.

The proposed object detection approach was evaluated on four subdatasets of the aforementioned Caltech-05 dataset: faces, motorbikes, airplanes and car sides. For the non-object images, the background subdataset was adopted for the four object categories. Each subdatasets was randomly split into two halves 15 times, one for training and the other for testing. As in the previously mentioned image classification task, the proposed model was used as a classifier for detecting objects via allocating the test image to the group (object or non-object) with the highest posterior probability according to the Bayesian decision rule.

Tab. 2 shows the detection accuracies along with the standard deviations for different approaches. Fig. 6 shows the distributions of the detection accuracies. These results indicate that the proposed approach performs better than the three other approaches and has the best detection accuracy, which confirms that the proposed method has better modeling capability than IBMM and IDMM, SMM in the case of non-Gaussian positive data.

V. CONCLUSIONS

We introduce a non-Gaussian mixture model with inverted Beta mixture components for modeling positive data. Compared to the conventional finite inverted Beta mixture model (IBMM), the proposed model has flexible model complexity, as the number of model parameters is variable and infinite. The recently proposed extended variational inference (EVI) framework is employed to perform parameter learning in the proposed model. Through two challenging applications, namely, image classification and object detection, we have demonstrated that the proposed model can provide good modeling and classification capabilities. Future work could be devoted to Bayesian estimation of the proposed model with the recently proposed stochastic variational inference, which would allow us to apply complicated Bayesian models to large-scale datasets.

REFERENCES

[1] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. Audio, Speech, Language Process.*, vol. 3, no. 1, pp. 72–83, Jan. 1995.
[2] G. McLachlan and D. Peel, *Finite Mixture Models*. New York, NY, USA: Wiley, 2000.
[3] N. Bouguila, D. Ziou, and J. Vaillancourt, "Unsupervised learning of a finite mixture model based on the Dirichlet distribution and its application," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1533–1543, Nov. 2004.
[4] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.
[5] Z. Ma and A. Leijon, "Bayesian estimation of beta mixture models with variational inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2160–2173, Nov. 2011.
[6] Z. Ma, P. K. Rana, J. Taghia, M. Flierl, and A. Leijon, "Bayesian estimation of Dirichlet mixture model with variational inference," *Pattern Recognit.*, vol. 47, no. 9, pp. 3143–3157, 2014.
[7] W. Fan, N. Bouguila, and D. Ziou, "Variational learning for finite Dirichlet mixture models and applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 5, pp. 762–774, May 2012.
[8] J. Taghia and A. Leijon, "Variational inference for Watson mixture model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1886–1900, Sep. 2015.
[9] T. P. Sahu, N. K. Nagwani, and S. Verma, "Multivariate Beta mixture model for automatic identification of topical authoritative users in community question answering sites," *IEEE Access*, vol. 4, pp. 5343–5355, 2016.
[10] X. Qiu, T. Jiang, S. Wu, and M. Hayes, "Physical layer authentication enhancement using a Gaussian mixture model," *IEEE Access*, vol. 6, pp. 53583–53592, 2018.
[11] M. Niknejad, H. Rabbani, and M. Babaie-Zadeh, "Image restoration using Gaussian mixture models with spatially constrained patch clustering," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3624–3636, Nov. 2015.
[12] N. Bouguila, "Hybrid generative/discriminative approaches for proportional data modeling and classification," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 12, pp. 2184–2202, Dec. 2012.
[13] T. Bdiri, N. Bouguila, and D. Ziou, "Variational Bayesian inference for infinite generalized inverted Dirichlet mixtures with feature selection and its application to clustering," *Appl. Intell.*, vol. 44, no. 3, pp. 507–525, 2016.
[14] Z. Ma, A. E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, "Variational Bayesian matrix factorization for bounded support data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 4, pp. 876–889, Apr. 2015.
[15] Z. Ma, J.-H. Xue, A. Leijon, Z.-H. Tan, Z. Yang, and J. Guo, "Decorrelation of neutral vector variables: Theory and applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 129–143, Jan. 2018.
[16] E. Epaillard and N. Bouguila, "Proportional data modeling with hidden Markov models based on generalized Dirichlet and Beta-Liouville mixtures applied to anomaly detection in public areas," *Pattern Recognit.*, vol. 55, pp. 125–136, Jul. 2016.
[17] Z. Ma, J. Xie, Y. Lai, J. Taghia, J.-H. Xue, and J. Guo, "Insights into multiple/single lower bound approximation for extended variational inference in non-Gaussian structured data modeling," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.
[18] Z. Ma, Y. Lai, W. B. Kleijn, Y.-Z. Song, L. Wang, and J. Guo, "Variational Bayesian learning for Dirichlet process mixture of inverted Dirichlet distributions in non-Gaussian image feature modeling," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 2, pp. 449–463, Feb. 2019.

[19] J. Ashburner and K. J. Friston, "Unified Segmentation," *NeuroImage*, vol. 26, no. 3, pp. 839–851, 2005.
[20] J.-D. Lee *et al.*, "MR image segmentation using a power transformation approach," *IEEE Trans. Med. Imag.*, vol. 28, no. 6, pp. 894–905, Jun. 2009.
[21] T. M. Nguyen and Q. M. J. Wu, "A nonsymmetric mixture model for unsupervised image segmentions," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 751–765, Apr. 2013.
[22] 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 Anal. Mach. Intell.*, vol. 34, no. 4, pp. 814–817, Apr. 2012.
[23] T. M. Nguyen and Q. M. J. Wu, "A non-parametric Bayesian model for bounded data," *Pattern Recognit.*, vol. 48, no. 6, pp. 2084–2095, Jun. 2015.
[24] G. J. McLachlan and D. Peel, *Finite Mixture Models*. New York, NY, USA: Wiley, 2000.
[25] A. Corduneanu and C. M. Bishop, "Variational Bayesian model selection for mixture distributions," in *Proc. 8th Int. Conf. Artif. Intell. Statist. (AISTATS)*, Orlando, FL, USA, Jan. 2001, pp. 1–8.
[26] Y. Lai, Y. Ping, B. Wang, J. Wang, and X. Zhang, "Variational Bayesian inference for finite inverted Dirichlet mixture model and its application to object detection," *Chin. J. Electron.*, vol. 27, no. 3, pp. 603–610, May 2018.
[27] M. Svensén and C. M. Bishop, "Robust Bayesian mixture modelling," *Neurocomputing*, vol. 64, nos. 1–4, pp. 235–252, 2005.
[28] T. Bdiri and N. Bouguila, "Positive vectors clustering using inverted Dirichlet finite mixture models," *Expert Syst. Appl.*, vol. 39, no. 2, pp. 1869–1882, 2012.
[29] U. Srinivas, H. S. Mousavi, V. Monga, A. Hattel, and B. Jayarao, "Simultaneous sparsity model for histopathological image representation and classification," *IEEE Trans. Med. Imag.*, vol. 33, no. 5, pp. 1163–1179, May 2014.
[30] H. Ma and Y. Yang, "Two specific multiple-level-set models for high-resolution remote-sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 3, pp. 558–561, Jul. 2009.
[31] L. Zhang, R. Hong, Y. Gao, R. Ji, Q. Dai, and X. Li, "Image categorization by learning a propagated graphlet path," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 3, pp. 674–685, Mar. 2016.
[32] Z. Ma *et al.*, "Fine-grained vehicle classification with channel max pooling modified CNNs," *IEEE Trans. Veh. Technol.*, to be published.
[33] M. T. Yang, R. K. Jhang, and J. S. Hou "Traffic flow estimation and vehicle-type classification using vision-based spatial-temporal profile analysis," *IET Comput. Vis.*, vol. 7, no. 5, pp. 394–404, Oct. 2013.
[34] C. Wojek, S. Walk, S. Roth, K. Schindler, and B. Schiele, "Monocular visual scene understanding: Understanding multi-object traffic scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 882–897, Apr. 2013.
[35] X. Li, H. Ma, X. Wang, and X. Zhang, "Traffic light recognition for complex scene with fusion detections," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 199–208, Jan. 2018.
[36] T. Goudas, C. Doukas, A. Chatziioannou, and I. Maglogiannis, "A collaborative biomedical image-mining framework: Application on the image analysis of microscopic kidney biopsies," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 1, pp. 82–91, Jan. 2013.
[37] M. X. Ribeiro, A. J. M. Traina, C. Traina, and P. M. Azevedo-Marques, "An association rule-based method to support medical image diagnosis with efficiency," *IEEE Trans. Multimedia*, vol. 10, no. 2, pp. 277–285, Feb. 2008.
[38] A. Gooya, H. Liao, K. Matsumiya, K. Masamune, Y. Masutani, and T. Dohi, "A variational method for geometric regularization of vascular segmentation in medical images," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1295–1312, Aug. 2008.
[39] Z. Shao, J. Cai, and Z. Wang, "Smart monitoring cameras driven intelligent processing to big surveillance video data," *IEEE Trans. Big Data*, vol. 4, no. 1, pp. 105–116, Mar. 2018.
[40] T. T. Zin, P. Tin, H. Hama, and T. Toriu, "Unattended object intelligent analyzer for consumer video surveillance," *IEEE Trans. Consum. Electron.*, vol. 57, no. 2, pp. 549–557, May 2011.
[41] J. S. Kim, D. H. Yeom, and Y. H. Joo, "Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems," *IEEE Trans. Consum. Electron.*, vol. 57, no. 3, pp. 1165–1170, Aug. 2011.
[42] Z. Ma, H. Yu, W. Chen, and J. Guo, "Short utterance based speech language identification in intelligent vehicles with time-scale modifications and deep bottleneck features," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 121–128, Jan. 2019.
[43] O. L. Junior, D. Delgado, V. Goncalves, and U. Nunes, "Trainable classifier-fusion schemes: An application to pedestrian detection," in *Proc. IEEE Int. Conf. Intell. Transp. Syst. (ICITS)*, St. Louis, MO, USA, Oct. 2009, pp. 1–6.
[44] M. Ge, Y. Xu, and R. Du, "An intelligent online monitoring and diagnostic system for manufacturing automation," *IEEE Trans. Autom. Sci. Eng.*, vol. 5, no. 1, pp. 127–139, Jan. 2008.
[45] M. Roopaei, P. Rad, and K.-K. R. Choo, "Cloud of things in smart agriculture: Intelligent irrigation monitoring by thermal imaging," *IEEE Cloud Comput.*, vol. 4, no. 1, pp. 10–15, Jan./Feb. 2017.
[46] R. Lu, X. Lin, Z. Shi, and X. S. Shen, "A lightweight conditional privacy-preservation protocol for vehicular traffic-monitoring systems," *IEEE Intell. Syst.*, vol. 28, no. 3, pp. 62–65, May/Jun. 2013.
[47] A. El-Mowafy and N. Kubo, "Integrity monitoring for positioning of intelligent transport systems using integrated RTK-GNSS, IMU and vehicle odometer," *IET Intell. Transp. Syst.*, vol. 12, no. 8, pp. 901–908, 2018.
[48] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 746–761, Apr. 2015.
[49] S. M. H. Anvar, W.-Y. Yau, and E. K. Teoh, "Multiview face detection and registration requiring minimal manual intervention," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2484–2497, Oct. 2013.
[50] R. Raghavendra, K. B. Raja, and C. Busch, "Presentation attack detection for face recognition using light field camera," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1060–1075, Mar. 2015.
[51] R. Usamentiaga, J. Molleda, D. F. Garcia, J. C. Granda, and J. L. Rendueles, "Temperature measurement of molten pig iron with slag characterization and detection using infrared computer vision," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 5, pp. 1149–1159, May 2012.
[52] Z. Tan, A. Jamdagni, X. He, P. Nanda, R. P. Liu, and J. Hu, "Detection of denial-of-service attacks based on computer vision techniques," *IEEE Trans. Comput.*, vol. 64, no. 9, pp. 2519–2533, Sep. 2015.

YUPING LAI received the Ph.D. degree in information security from the Beijing University of Posts and Telecommunications, Beijing, China, in 2014. He has been an Associate Professor with the North China University of Technology, China, since 2014. His research interests include information security, computer vision, pattern recognition, machine learning, and data mining.

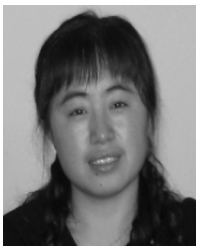
XIU MA received the bachelor's degree in information security from the North China University of Technology, Beijing, in 2017, where she is currently pursuing the master's degree with the North China University of Technology. She majors in pattern recognition and machine learning.

YANPING XU received the Ph.D. degree in computer science from the Beijing University of Posts and Telecommunications, China, in 2017. She is currently a Lecturer with Hangzhou Dianzi University. Her research interests include cyberspace security, industrial control security, intrusion detection, machine learning, and data mining.

YONGFA LING received the Ph.D. degree in automatic control from the Kunming University of Science and Technology, China, in 2004. He did his Postdoctoral research at Xi'an Jiaotong University, Xi'an, China, in 2007. He is currently a Professor with Hezhou University, China. His research interests include wireless sensor networks, compressed sensing, MAC mechanism and system simulation, MPLS traffic engineering, conflict resolution algorithm, and priority control technology.

CHUNLAI DU received the Ph.D. degree in computer architecture from the Harbin Institute of Technology, China, in 2009. He is currently an Associate Professor with the North China University of Technology, China. His research interests include malware detection, vulnerability exploitation, and network security.

JIANHE DU received the Ph.D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2015. He is currently a Lecturer with the School of Information and Engineering, Communication University of China, Beijing. His research interests mainly include channel estimation, symbol detection, multiple-antenna systems, MIMO relay systems, and tensor-based signal processing applied to wireless communication.



YONGMEI ZHANG received the B.S. degree from Peking University, Beijing, China, in 1990, and the M.Sc. and Ph.D. degrees from the North University of China, Shanxi, China. She did her Postdoctoral research at Beihang University, Beijing, in 2008. In 2012, she was a Visiting Scholar with Peking University. She is currently a Professor with the North China University of Technology. Her research interests include artificial intelligence, pattern recognition, and image processing. She is a member of CCF and ACM.

YUAN PING received the B.S. degree in electronics and information engineering from Southwest Normal University, in 2003, the M.S. degree in mathematics from Henan University, in 2008, and the Ph.D. degree in information security from the Beijing University of Posts and Telecommunications, in 2012. He was a Visiting Scholar with the Department of Computing Science, University of Alberta. He is currently an Associate Professor with Xuchang University and a Visiting Scholar with the School of Computing and Informatics, University of Louisiana at Lafayette. His research interests include machine learning, public-key cryptography, data privacy and security, and cloud and edge computing.