2015 IEEE 14th International Conference on Machine Learning and Applications

Human Action Recognition using Accelerated Variational Learning of Infinite Dirichlet Mixture Models

Wentao Fan*, Hassen Sallay†, Nizar Bouguila‡ and Ji-Xiang Du*
*Department of Computer Science and Technology, Huaqiao University, Xiamen, China
Email: fwt@hqu.edu.cn; jxdu@hqu.edu.cn
†College of Computer and Information Systems, Umm Al-Qura University, Saudi Arabia
Email: hmsallay@uqu.edu.sa
‡Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada
Email: nizar.bouguila@concordia.ca

***Abstract*—Exploiting Dirichlet process mixture models (also known as infinite mixture models) to model visual and textual data is now standard weapon in the arsenal of machine learning. This paper proposes a new accelerated variational inference approach to learn Dirichlet process mixture models with Dirichlet distributions. The choice of using Dirichlet distribution as the basic distribution is mainly due to its flexibility for modeling proportional data. Indeed, this kind of data is naturally generated by several applications involving the representation of texts, images and videos using the bag-of-words (or "visual words" in the case of images and videos) approach. The potential of the developed learning framework is shown using a challenging real application namely human action recognition in videos.**

***Keywords*-Mixture models; clustering; Dirichlet process; nonparametric Bayesian; variational inference; human action recognition**

I. Introduction

The applications of finite mixture models have widened dramatically in the past few years [1]. In particular, models with symmetric distributions (e.g. Gaussian, generalized Gaussian) have received a lot of attention [2]. However, in many real-life applications data have asymmetric non-Gaussian forms as we have shown in many of our recent research works (see, for instance, [3], [4], [5], [6]). This is the case of proportional data for which the finite Dirichlet mixture model has been shown to be clearly an excellent choice in many image processing and computer vision applications. Examples include content-based images categorization and retrieval [4], [5] and shadows detection in images [7]. As compared to the finite Dirichlet mixture, the infinite one which is based on Dirichlet process mixture models [8] has been shown to provide better modeling results and generalization capabilities [9], [10], [11]. Indeed, infinite mixture provides better generalization capability by allowing model's complexity to increase as new data arrive. A crucial problem when deploying infinite mixture models is the learning of the parameters. Markov chain Monte Carlo (MCMC) [12] sampling techniques have been widely used to learn infinite mixtures in general and the infinite Dirichlet mixture in particular [9]. However, a main limitation with these techniques is that they are computationally expensive. A different approach that overcomes this problem considers deterministic approximation based on variational learning [13], [14]. In [15], a kd-tree structure was adopted in the variational inference for learning Dirichlet process mixtures with exponential family, in order to improve the computational efficiency. It is noteworthy that although Dirichlet distribution belongs to the exponential family, the general variational solution to the Dirichlet process mixtures with exponential family proposed in [15] can not be performed directly in our case. This is due to the fact that the formal conjugate prior for the Dirichlet distribution is intractable, mainly because of the difficulty to evaluate the corresponding normalization coefficient, and cannot be applied for variational inference directly. This problem is tackled by introducing an approximate prior which generates an explicit and tractable solution as described later in this paper.

The main contribution of our work is to propose an accelerated variational approach for infinite Dirichlet mixtures learning based on the framework developed in [15]. Moreover, compared with [15] in which Gaussian mixtures are used for all their experiments, our work is based on Dirichlet mixtures. In contrast to Gaussian distribution which only contains symmetric modes, the Dirichlet distribution may have multiple symmetric and asymmetric modes. Therefore, Dirichlet mixtures may provide more flexibility than Gaussian mixtures. The proposed approach is easy to implement and provides effective and accurate learning as we demonstrate via extensive simulations that concern the challenging application of human activity recognition.

The paper is organized as follows. Section 2 introduces the infinite Dirichlet mixture model using stick-breaking representation and proposes the learning framework. Section 3 presents the results of applying the proposed model on the challenging problem of human activity recognition. Conclusions are reported in Section 4.

978-1-5090-0287-0/15 $31.00 © 2015 IEEE
DOI 10.1109/ICMLA.2015.68

II. The Infinite Dirichlet Mixture Model With Stick-breaking Representation

A finite mixture of Dirichlet distributions with M components is given by [16], [17]

$$p(\vec{X}|\vec{\pi},\vec{\alpha}) = \sum_{j=1}^{M} \pi_j \mathrm{Dir}(\vec{X}|\vec{\alpha}_j) \ , \tag{1}$$

where $\vec{\pi} = (\pi_1, \ldots, \pi_M)$ represents the mixing coefficients which are positive and sum to one. $\mathrm{Dir}(\vec{X}|\vec{\alpha}_j)$ is the Dirichlet distribution of component j with its own positive parameters $\vec{\alpha}_j = (\alpha_{j1}, \ldots, \alpha_{jD})$ and is defined by

$$\mathrm{Dir}(\vec{X}|\vec{\alpha}_j) = \frac{\Gamma(\sum_{l=1}^{D}\alpha_{jl})}{\prod_{l=1}^{D}\Gamma(\alpha_{jl})} \prod_{l=1}^{D} X_l^{\alpha_{jl}-1} \tag{2}$$

where $\vec{X} = (X_1, \ldots, X_D)$ and $\sum_{l=1}^{D} X_l = 1$, $X_l > 0$ for $l = 1, \ldots, D$.

A crucial issue when using mixture models is the model complexity (i.e. model structure or number of mixture components) determination problem. Indeed, it is important to estimate the number of clusters M that best describes the data without over-fitting or under-fitting it. This difficulty can be tackled elegantly by using a nonparametric Bayesian framework namely Dirichlet process (DP) [8] mixture model by assuming that M is infinite [18]. In our work, we adopt the stick-breaking representation [19] to build the DP. Specifically, G is a DP with base distribution H and scaling parameter φ, denoted as $G \sim \mathrm{DP}(\varphi, H)$, if the following conditions are satisfied

$$\begin{aligned} &\lambda \sim \mathrm{Beta}(1,\varphi) \ , \quad \Omega_j \sim H \\ &\pi_j = \lambda_j \prod_{s=1}^{j-1}(1-\lambda_s) \ , \quad G = \sum_{j=1}^{\infty} \pi_j \delta_{\Omega_j} \end{aligned} \tag{3}$$

where δ_{Ω_j} denotes the Dirac delta measure centered at Ω_j.

The DP mixture of Dirichlet distributions with stick-breaking construction is defined as follows. Assuming that we have obtained a data set $\mathcal{X} = (\vec{X}_1, \ldots, \vec{X}_N)$ which is distributed according to a Dirichlet mixture model with an infinite number of components. We then introduce a vector $\vec{Z} = (Z_1, \ldots, Z_N)$ as the mixture component assignment variable, so that each element Z_i takes an integer value j denoting the component from which $\vec{X}_i$ is drawn. The probability of Z is defined in terms of the mixing coefficients π_j as

$$p(\vec{Z}|\pi) = \prod_{i=1}^{N}\prod_{j=1}^{\infty} \pi_j^{\mathbf{1}[Z_i=j]} \tag{4}$$

where $\mathbf{1}[\cdot]$ is an indicator function which has the value 1 when $Z_i = j$ and 0 otherwise. Since π_j is a function of $\vec{\lambda}$ as shown in (3), the probability of $\vec{Z}$ can also be written as

$$p(\vec{Z}|\vec{\lambda}) = \prod_{i=1}^{N}\prod_{j=1}^{\infty}\left[\lambda_j \prod_{s=1}^{j-1}(1-\lambda_s)\right]^{\mathbf{1}[Z_i=j]} \tag{5}$$

According to the stick-breaking representation (3), the prior distribution of $\vec{\lambda}$ is a specific Beta distribution in the following form

$$p(\vec{\lambda}|\vec{\varphi}) = \prod_{j=1}^{\infty} \mathrm{Beta}(1,\varphi_j) = \prod_{j=1}^{\infty} \varphi_j (1-\lambda_j)^{\varphi_j - 1} \tag{6}$$

Then the likelihood function of $\mathcal{X}$ given $\vec{Z}$ and $\vec{\alpha}$ is equivalent to the distribution of $\vec{X}_i$ conditioned on the Z_ith component in the mixture as

$$\begin{aligned} p(\mathcal{X}|\vec{Z},\vec{\alpha}) &= \prod_{i=1}^{N} p(\vec{X}_i|\vec{\alpha}_{Z_i}) \\ &= \prod_{j=1}^{\infty}\left\{\left[\frac{\Gamma(\sum_{l=1}^{D}\alpha_{jl})}{\prod_{l=1}^{D}\Gamma(\alpha_{jl})}\right]^{n_j} \prod_{l=1}^{D}\Big(\prod_{i=1}^{N} X_{il}^{\mathbf{1}[Z_i=j]}\Big)^{\alpha_{jl}-1}\right\} \end{aligned} \tag{7}$$

where n_j denotes the number of observations belonging to class j and is defined as $n_j = \sum_{i=1}^{N} \mathbf{1}[Z_i = j]$.

In our Bayesian framework, a prior distribution over $\vec{\alpha}$ needs to be introduced. Since the formal conjugate prior for the Dirichlet distribution is intractable, it cannot be applied for the variational inference directly. In this case, a Gamma distribution is adopted to approximate the conjugate prior over $\vec{\alpha}$ by assuming that parameters $\{\alpha_{jl}\}$ are statistically independent

$$p(\vec{\alpha}) = \prod_{j=1}^{\infty}\prod_{l=1}^{D} \frac{v_{jl}^{u_{jl}}}{\Gamma(u_{jl})} \alpha_{jl}^{u_{jl}-1} e^{-v_{jl}\alpha_{jl}} \tag{8}$$

where u_{jl} and v_{jl} are positive hyperparameters.

A. Model Learning

Recently, several methods have been proposed for learning infinite Dirichlet mixture models, such as the stochastic approach [9] via Markov chain Monte Carlo (MCMC) and the deterministic approach [14] through variational inference [20], [13]. Although both the MCMC and the variational inference methods are able to learn infinite Dirichlet mixture model effectively, they would suffer when encountering huge volume of data (e.g., millions of data instances). This problem can be tackled using an accelerated version of variational inference method as proposed in [15]. In this section, we propose an accelerated variational inference method based on kd-tree structure for learning infinite Dirichlet mixture models.

The main goal of variational inference is to find an approximation $q(\Theta)$ for the posterior distribution $p(\Theta|\mathcal{X})$, where $\Theta = \{\vec{Z}, \vec{\alpha}, \vec{\lambda}\}$ is the set of random variables associated with our infinite Dirichlet mixture model. By adopting the truncated stick-breaking representation and the factorization assumption, we can obtain

$$q(\Theta) = q(\vec{Z})q(\vec{\alpha})q(\vec{\lambda}) = \left[\prod_{i=1}^{N} q(Z_i)\right]\left[\prod_{j=1}^{\infty} q(\lambda_j)\prod_{l=1}^{D} q(\alpha_{jl})\right] \tag{9}$$

A common trick for learning infinite mixture models is the adoption of truncation techniques [21], [14] to truncate the number of mixture components of variational posteriors into a finite level. However, as indicated in [15], explicitly truncating variational posteriors may generate undesirable consequence that the approximating variational families are not nested. In order to tackle this problem, we follow the setting as stated in [15]: the number of mixture components for variational posteriors remain infinite, but the variational parameters of all models are tied after a specific level M. More specifically, if a component is associated with index $j > M$, then $q(\lambda_j)$ and $q(\vec{\alpha}_j)$ are equal to their corresponding priors. In order to obtain the variational posteriors with respect of the parameter tying assumption for $j > M$, the following free energy is required to be minimized

$$F = \sum_{i=1}^{N} \big\langle \ln \frac{q(Z_i)}{p(Z_i|\vec{\lambda})p(\vec{X}_i|\vec{\alpha}_{Z_i})} \big\rangle + \sum_{j=1}^{M} \Big[\sum_{l=1}^{D} \big\langle \ln \frac{q(\alpha_{jl})}{p(\alpha_{jl})} \big\rangle + \big\langle \frac{q(\lambda_j)}{p(\lambda_j)} \big\rangle \Big] \tag{10}$$

where $\langle \cdot \rangle$ represents the corresponding expected value. Thus, we can obtain the variational solutions as follows

$$q(\vec{Z}) = \prod_{i=1}^{N} \prod_{j=1}^{M} r_{ij}^{\mathbf{1}[Z_i=j]} \tag{11}$$

$$q(\vec{\alpha}) = \prod_{j=1}^{M} \prod_{l=1}^{D} \mathcal{G}(\alpha_{jl}|u_{jl}^*, v_{jl}^*) \tag{12}$$

$$q(\vec{\lambda}) = \prod_{j=1}^{M} \mathrm{Beta}(\lambda_j|a_j, b_j) \tag{13}$$

where the associated hyperparameters are calculated by

$$r_{ij} = \frac{\exp(\rho_{ij})}{\sum_{j=1}^{\infty} \exp(\rho_{ij})} \tag{14}$$

$$\rho_{ij} = \widetilde{\mathcal{R}}_j + \sum_{l=1}^{D} (\bar{\alpha}_{jl} - 1) \ln X_{il} + \big\langle \ln \lambda_j \big\rangle + \sum_{s=1}^{j-1} \big\langle \ln(1-\lambda_s) \big\rangle \tag{15}$$

$$\sum_{j=M+1}^{\infty} \exp(\rho_{ij}) = \frac{\exp(\rho_{i,M+1})}{1 - \exp[\psi(\varphi_j) - \psi(1+\varphi_j)]} \tag{16}$$

$$u_{jl}^* = u_{jl} + \sum_{i=1}^{N} r_{ij} \bar{\alpha}_{jl} \Big[\sum_{s \neq l}^{D} \Psi'(\sum_{s=1}^{D} \bar{\alpha}_{js}) \bar{\alpha}_{js} (\big\langle \ln \alpha_{js} \big\rangle - \ln \bar{\alpha}_{js}) + \Psi(\sum_{l=1}^{D} \bar{\alpha}_{jl}) - \Psi(\bar{\alpha}_{jl}) \Big] \tag{17}$$

$$v_{jl}^* = v_{jl} - \sum_{i=1}^{N} r_{ij} \ln X_{il} \tag{18}$$

$$a_j = 1 + \sum_{i=1}^{N} \big\langle Z_i = j \big\rangle \tag{19}$$

$$b_j = \varphi_j + \sum_{i=1}^{N} \sum_{s=j+1}^{\infty} \big\langle Z_i = s \big\rangle \tag{20}$$

where $\Psi(\cdot)$ and $\Psi'(\cdot)$ are the digamma and trigamma functions, respectively. $\widetilde{\mathcal{R}}_j$ in (15) is the lower bound of $\mathcal{R}_j = \big\langle \ln \frac{\Gamma(\sum_{l=1}^{D} \alpha_{jl})}{\prod_{l=1}^{D} \Gamma(\alpha_{jl})} \big\rangle$ obtained using a second-order Taylor expansion. It is noteworthy that since the parameter tying assumption for $j > M$ is adopted, $\sum_{s=j+1}^{\infty} \big\langle Z_i = s \big\rangle$ in (20) can be calculated from (14) and (16). The expected values in the above formulas are given by

$$\big\langle Z_i = j \big\rangle = r_{ij} \ , \qquad \bar{\alpha}_{jl} = \langle \alpha_{jl} \rangle = \frac{u_{jl}^*}{v_{jl}^*} \tag{21}$$

$$\big\langle \ln \alpha_{jl} \big\rangle = \Psi(u_{jl}^*) - \ln v_{jl}^* \tag{22}$$

$$\big\langle \ln \lambda_j \big\rangle = \Psi(a_j) - \Psi(a_j + b_j) \tag{23}$$

$$\big\langle \ln(1 - \lambda_j) \big\rangle = \Psi(b_j) - \Psi(a_j + b_j) \tag{24}$$

Next, the above variational inference procedure is expanded into an accelerate version as proposed in [15] though a kd-tree structure [22]. Assume that we have stored the data set $\mathcal{X}$ in a kd-tree with the constraint that all data points x_i in outer node T share the same responsibility $q(Z_i) \equiv q(Z_T)$. Under this constraint, the variational solutions with kd-tree structure can be obtained as follows

$$\big\langle Z_T = j \big\rangle = \frac{\exp(\rho_{Tj})}{\sum_{j=1}^{\infty} \exp(\rho_{Tj})} \tag{25}$$

$$\rho_{Tj} = \widetilde{\mathcal{R}}_j + \sum_{l=1}^{D} (\bar{\alpha}_{jl} - 1) \ln \langle X_l \rangle_T + \big\langle \ln \lambda_j \big\rangle + \sum_{s=1}^{j-1} \big\langle \ln(1-\lambda_s) \big\rangle \tag{26}$$

$$u_{jl}^* = u_{jl} + \sum_{T} |n_T| \big\langle Z_T = j \big\rangle \bar{\alpha}_{jl} \Big[\Psi(\sum_{l=1}^{D} \bar{\alpha}_{jl}) - \Psi(\bar{\alpha}_{jl}) + \sum_{s \neq l}^{D} \Psi'(\sum_{s=1}^{D} \bar{\alpha}_{js}) \bar{\alpha}_{js} (\big\langle \ln \alpha_{js} \big\rangle - \ln \bar{\alpha}_{js}) \Big] \tag{27}$$

$$v_{jl}^* = v_{jl} - \sum_{T} |n_T| \big\langle Z_T = j \big\rangle \ln \langle X_l \rangle_T \tag{28}$$

$$a_j = 1 + \sum_{T} |n_T| \big\langle Z_T = j \big\rangle \tag{29}$$

$$b_j = \varphi_j + \sum_{T} |n_T| \sum_{s=j+1}^{\infty} \big\langle Z_T = s \big\rangle \tag{30}$$

where $\langle X_l \rangle_T$ represents average over all data $\vec{X}_i$ contained in node T, $|n_T|$ denotes the number of data in node T. In contrast with the variational learning algorithm without using kd-tree where $O(MN)$ cost is required for each update cycle, the computational efficiency is significantly improved ($O(M|T|)$) when the kd-tree structure is adopted. The complete learning process with kd-tree structure is summarized in Algorithm 1. Similar as in [15], components

are reordered every one update cycle and the kd-tree is expanded every three cycles. The expansion of the kd-tree is controlled through the relative change of $\langle Z_T \rangle$ between a node and its children.

Algorithm 1

1: **Input:** A date set $\mathcal{X}$ which is already stored in a kd-tree structure
2: **Output:** $\{\vec{u}_j^*, \vec{v}_j^*, a_j, b_j\}_{j=1}^M, M$
3: {Initialization}
4: Set $M = 1$. Expand the kd-tree to some initial level (e.g. five)
5: Sample a number of 'candidate' components c according to size $\sum_T |n_T| \langle Z_T = c \rangle$
6: **for** each candidate c **do**
7: Expand one-level deeper the outer nodes of the kd-tree that assign to c the highest responsibility $\langle Z_T = c \rangle$ among all components
8: Split c in to two components, c_1 and c_2, through the bisector of its principal component. Initialize the responsibilities $\langle Z_T = c_1 \rangle$ and $\langle Z_T = c_2 \rangle$
9: Update ρ_{Tc_1}, $\vec{u}_{c_1}^*$, $\vec{v}_{c_1}^*$, a_{c_1}, b_{c_1} and ρ_{Tc_2}, $\vec{u}_{c_2}^*$, $\vec{v}_{c_2}^*$, a_{c_2}, b_{c_2} for new components c_1 and c_2
10: **end for**
11: Update ρ_{Tj}, $\vec{u}_j^*$, $\vec{v}_j^*$, a_j, b_j for all $j \leq M + 1$, while expanding the kd-tree and reordering components
12: **if** $F_M - F_{M+1} < threhold$ **then**
13: Stop
14: **else**
15: set $M = M + 1$ and go to step 5
16: **end if**
17: **return** $\{\vec{u}_j^*, \vec{v}_j^*, a_j, b_j\}_{j=1}^M$ and M

III. Experiments on Human Action Videos Recognition

Due to the improvement of digital technologies, the availability of digital videos is rapidly increasing. With thousands of videos on hand, grouping them according to their contents is highly demanded and is also a critical task which can be used to organize, summarize and retrieve this massive amount of data. In this experiment, we focus on developing a novel statistical approach for recognizing human action videos based on local spatio-temporal features with bag of visual words representation and the proposed infinite Dirichlet mixture model (*InDMM*).

A. Methodology and Data Set

The methodology that we have adopted for recognizing human actions in videos can be summarized as follows. First, local spatio-temporal features were extracted from each video sequence from their detected space-time interest points. In our case, we use the Harris3D detector [23] to obtain the HOG/HOF feature descriptors [24]. Specifically, for each clip, a set of 3D Harris corners is detected, and a local descriptor is computed as a concatenation of the HOG and HOF around the 3D Harris corner. Next, a visual vocabulary is constructed by quantizing these spatio-temporal features into visual words using K-means algorithm and each video is then represented as a frequency histogram over the visual words. Then, the pLSA model [25] is adopted as a dimension reduction method to represent each video sequence as a D-dimensional proportional vector, where D is the number of latent aspects. Finally, the proposed *InDMM* is applied to recognizing human actions by assigning the video sequence to the action category which has the highest posterior probability according to Bayes' decision rule.

Our experiments were conducted on the HMDB51 database [26][1], which is a large human motion database containing 51 action categories with about 7,000 clips in total. HMDB51 database was collected from various sources, such as movies, the Prelinger archive, YouTube and Google videos.

Figure 1. Examples of frames of different human actions from video sequences in the HMDB51 human action data set.

B. Results

In our first experiment, we used a subset of the HMDB51 database which includes 10 actions: catch, eat, hit, hug, jump, kick, kiss, push, sit and walk. Each action category contains 100 video sequences and results in 1,000 video sequences in total. Some examples of frames are displayed in Fig. 1. We randomly divided this test data set into two halves: one for constructing the visual vocabulary and the other for testing. In this experiment, the visual vocabulary was built by setting the number of clusters in the K-Means algorithm (i.e. number of visual words) to 1200, as explained in the previous section. The pLSA model was applied by considering 50 aspects and each video sequence in the tested data set was then represented by a 50-dimensional vector of proportions. We evaluated the recognition performance of the proposed algorithm by running it 20 times. The average recognition accuracy for each action category is shown in Fig. 2 and the total accuracy is 69.3%.

We have applied two state-of-the-art approaches with the same experimental settings for comparison: the infinite

[1] http://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database

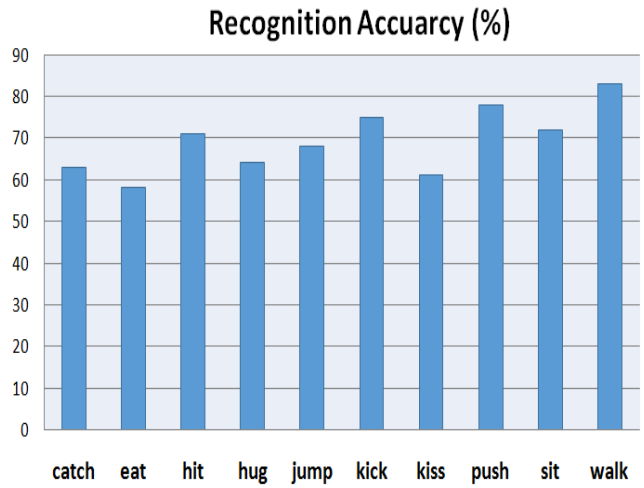

Figure 2. Average recognition accuracy for each action category using the proposed *InDMM*.

Gaussian mixture model with kd-tree structure as proposed in [15], and the approach as described in [26] where SVM with an RBF kernel is adopted for classification. The average recognition accuracy obtained by each approach is shown in Table I. According to this table, it is obvious that *InDMM* outperformed the other two approaches in terms of the highest recognition accuracy (69.3%). A student's t-test shows that the improvement is statistically significant (p-values between 0.032 and 0.041).

Table I
THE AVERAGE RECOGNITION ACCURACY WITH THE STANDARD DEVIATIONS VIA DIFFERENT ALGORITHMS IN 20 RUNS.

Method	10 Actions	51 Actions
InDMM	69.3 $\pm$1.7	27.7 $\pm$1.2
Kurihara *et al.* [15]	65.8 $\pm$ 1.5	24.5 $\pm$1.1
Kuehne *et al.* [26]	62.7 $\pm$ 1.7	22.1 $\pm$0.8

Table II
THE AVERAGE COMPUTATIONAL RUN TIME (IN HOURS) VIA DIFFERENT ALGORITHMS IN 20 RUNS.

Method	10 Actions	51 Actions
InDMM	0.36	2.52
Fan *et al.* [14]	1.08	6.94

Additionally, we conducted the experiment using the whole HMDB51 database with 51 action categories. The corresponding average recognition results are demonstrated in Table I. Based on this table, the proposed *InDMM* algorithm also provided the best recognition performance with the highest recognition accuracy (27.7%). According to the student's t-test, the difference between the *InDMM* and the other two approaches is statistically significant (p-values between 0.027 and 0.046).

In order to demonstrate the advantages of using the accelerated variational inference approach, we compare the computational costs (in terms of computational run time) obtained by the proposed *InDMM* and by the conventional variational learning approach without kd-tree structure as proposed in [14] for recognizing human actions. The corresponding results are shown in Table II. It is worth mentioning that the results reported in this table only with respect to the classification step. The run time for extracting features and constructing visual vocabulary was not included in this table. According to this table, it is obvious that the proposed *InDMM* gains much more computational efficiency than the variational learning approach without using kd-tree stricture.

IV. CONCLUSION

We adopted an accelerated variational approach to learn infinite Dirichlet mixture models. This approach is easy to implement and offers more generalization capabilities. Experiments that concern the challenging task of human actions recognition show that this approach provides good modeling and clustering results. Future works could be devoted to the implementation of the proposed inference framework in the case of the infinite generalized Dirichlet mixture or to extend the proposed inference technique to deal with dynamic settings.

ACKNOWLEDGMENT

The completion of this research was made possible thanks to the Natural Sciences and Engineering Research Council of Canada (NSERC), the Scientific Research Funds of Huaqiao University (600005-Z15Y0016), the Grant of the National Science Foundation of China (No. 61175121), the Program for New Century Excellent Talents in University (No.NCET-10-0117), the Grant of the National Science Foundation of Fujian Province (No.2013J06014), the Promotion Program for Young and Middle-aged Teacher in Science and Technology Research of Huaqiao University (No.ZQN-YX108). The second author would like to thank King Abdulaziz City for Science and Technology (KACST), Kingdom of Saudi Arabia, for their funding support under grant number 11-INF1787-08.

REFERENCES

[1] G. McLachlan and D. Peel, *Finite Mixture Models*. New York: Wiley, 2000.

[2] T. Elguebaly and N. Bouguila, "Bayesian learning of generalized gaussian mixture models on biomedical images," in *Artificial Neural Networks in Pattern Recognition, 4th IAPR TC3 Workshop, ANNPR 2010, Cairo, Egypt, April 11-13, 2010. Proceedings* (F. Schwenker and N. E. Gayar, eds.), vol. 5998 of *Lecture Notes in Computer Science*, pp. 207–218, Springer, 2010.

[3] N. Bouguila and D. Ziou, "A powerful finite mixture model based on the generalized dirichlet distribution: Unsupervised learning and applications," in *Proc. of the 17th International Conference on Pattern Recognition (ICPR)*, pp. 280–283, 2004.

[4] N. Bouguila and D. Ziou, "Mml-based approach for finite dirichlet mixture estimation and selection," in *Machine Learning and Data Mining in Pattern Recognition, 4th International Conference, MLDM 2005, Leipzig, Germany, July 9-11, 2005, Proceedings* (P. Perner and A. Imiya, eds.), vol. 3587 of *Lecture Notes in Computer Science*, pp. 42–51, Springer, 2005.

[5] N. Bouguila and D. Ziou, "On fitting finite dirichlet mixture using ECM and MML," in *Pattern Recognition and Data Mining, Third International Conference on Advances in Pattern Recognition, ICAPR 2005, Bath, UK, August 22-25, 2005, Proceedings, Part I* (S. Singh, M. Singh, C. Apté, and P. Perner, eds.), vol. 3686 of *Lecture Notes in Computer Science*, pp. 172–182, Springer, 2005.

[6] W. Fan, N. Bouguila, and D. Ziou, "A variational statistical framework for object detection," in *Neural Information Processing - 18th International Conference, ICONIP 2011, Shanghai, China, November 13-17, 2011, Proceedings, Part II* (B. Lu, L. Zhang, and J. T. Kwok, eds.), vol. 7063 of *Lecture Notes in Computer Science*, pp. 276–283, Springer, 2011.

[7] N. Bouguila and D. Ziou, "A probabilistic approach for shadows modeling and detection," in *Proc. of the International Conference on Image Processing (ICIP)*, pp. 329–332, 2005.

[8] R. M. Korwar and M. Hollander, "Contributions to the theory of Dirichlet processes," *The Annals of Probability*, vol. 1, pp. 705–711, 1973.

[9] N. Bouguila and D. Ziou, "A Dirichlet process mixture of Dirichlet distributions for classification and prediction," in *Proc. of the IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 297–302, 2008.

[10] W. Fan and N. Bouguila, "Infinite dirichlet mixture model and its application via variational bayes," in *10th International Conference on Machine Learning and Applications and Workshops, ICMLA 2011, Honolulu, Hawaii, USA, December 18-21, 2011. Volume 1: Main Conference* (X. Chen, T. S. Dillon, H. Ishbuchi, J. Pei, H. Wang, and M. A. Wani, eds.), pp. 129–132, IEEE Computer Society, 2011.

[11] W. Fan and N. Bouguila, "Online variational finite dirichlet mixture model and its applications," in *11th International Conference on Information Science, Signal Processing and their Applications, ISSPA 2012, Montreal, QC, Canada, July 2-5, 2012*, pp. 448–453, IEEE, 2012.

[12] C. Robert and G. Casella, *Monte Carlo Statistical Methods*. Springer-Verlag, 1999.

[13] A. Corduneanu and C. M. Bishop, "Variational Bayesian model selection for mixture distributions," in *Proc. of the 8th International Conference on Artificial Intelligence and Statistics (AISTAT)*, pp. 27–34, 2001.

[14] W. Fan and N. Bouguila, "Variational learning for Dirichlet process mixtures of Dirichlet distributions and applications," *Multimedia Tools and Applications*, vol. 70, no. 3, pp. 1685–1702, 2014.

[15] K. Kurihara, M. Welling, and N. Vlassis, "Accelerated variational dirichlet process mixtures," in *Proc. of Advances in Neural Information Processing Systems (NIPS)*, 2006.

[16] N. Bouguila, D. Ziou, and J. Vaillancourt, "Unsupervised learning of a finite mixture model based on the Dirichlet distribution and its application," *IEEE Trans. on Image Processing*, vol. 13, no. 11, pp. 1533–1543, 2004.

[17] W. Fan, N. Bouguila, and D. Ziou, "Variational learning for finite dirichlet mixture models and applications," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 23, no. 5, pp. 762–774, 2012.

[18] C. E. Rasmussen, "The infinite Gaussian mixture model," in *Proc. of Neural Information Processing Systems (NIPS)*, pp. 554–560, 2000.

[19] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica*, vol. 4, pp. 639–650, 1994.

[20] H. Attias, "A variational Bayes framework for graphical models," in *Proc. of Advances in Neural Information Processing Systems (NIPS)*, pp. 209–215, 1999.

[21] D. Blei and M. Jordan, "Variational inference for Dirichlet process mixtures," *Bayesian Analysis*, vol. 1, pp. 121–144, 2005.

[22] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, pp. 509–517, Sept. 1975.

[23] I. Laptev, "On space-time interest points," *International Journal of Computer Vision*, vol. 64, no. 2/3, pp. 107–123, 2005.

[24] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, "Learning realistic human actions from movies," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2008.

[25] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Machine Learning*, vol. 42, no. 1/2, pp. 177–196, 2001.

[26] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "HMDB: a large video database for human motion recognition," in *Proc. of the International Conference on Computer Vision (ICCV)*, pp. 2556 – 2563, 2011.