Proc. of International Conference
on Machine Learning and Applications, 2012.

Regularized Probabilistic Latent Semantic Analysis with Continuous Observations

Hao Zhang, Richard Edwards, and Lynne Parker
Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville, Tennessee 37996
{*haozhang, redwar15, leparker*}*@utk.edu*

***Abstract*—Probabilistic latent semantic analysis (PLSA) has been widely used in the machine learning community. However, the original PLSAs are not capable of modeling real-valued observations and usually have severe problems with overfitting. To address both issues, we propose a novel, regularized Gaussian PLSA (RG-PLSA) model that combines Gaussian PLSAs and hierarchical Gaussian mixture models (HGMM). We evaluate our model on supervised human action recognition tasks, using two publicly available datasets. Average classification accuracies of 97.69% and 93.72% are achieved on the Weizmann and KTH Action Datasets, respectively, which demonstrate that the RG-PLSA model outperforms Gaussian PLSAs and HGMMs, and is comparable to the state of the art.**

***Keywords*-PLSA; Gaussian mixture models; continuous features; human action recognition**

I. Introduction

Probabilistic latent semantic analysis (PLSA) [1] is extensively studied in the machine learning community and was originally proposed to categorize large collections of documents into a set of pre-defined topics. In this work, we focus on the task of human action recognition [2]. Our objective is to automatically assign each video with an action chosen from a fixed number of predefined action categories. Human action recognition is a challenging task, due to substantial variations within an action, especially when performed by different humans. Similarly, the same action by the same human can be performed at different speeds with different poses, giving rise to temporal variations.

The original PLSA model is based on the "bag-of-words" assumption, in which features are assumed to be discrete. However, features that are used to classify human actions are usually continuously distributed in some high dimensional space. Moreover, the PLSA model assumes that each observation has a different distribution over the pre-defined categories. Since the number of model parameters increases linearly with the number of observations, the PLSA model usually suffers from a severe overfitting problem. For example, in human action recognition, the PLSA model assigns a different distribution over the possible action categories to each observation. However, it is highly possible that a human performs the same action in consecutive observations. In this case, PLSA models overemphasize the action variations within the same category, which can lead to overfitting.

Gaussian mixture models (GMMs) can model continuous features and provide a way to prevent overfitting. In GMMs, observations within the same category are assumed to have the same per-observation category distribution. GMMs can achieve reasonable accuracy when modeling data generated from Gaussian distributions. However, GMMs are limited in their expressive power. These models can easily underfit if the true distributions are complex. For example, in human action recognition, GMMs assume that actions within the same category have the same distribution, and they ignore variations within actions performed by different people, or by the same person with different poses.

We introduce a new latent variable graphical model, called the *Regularized Gaussian PLSA* (RG-PLSA) model, which applies a regularization term to combine the advantages of the GMM model and the PLSA model. Our contributions are two-fold: 1) the RG-PLSA model extends the original PLSA model to support continuous real-valued observations; 2) regularization is used to reduce model complexity, which simultaneously prevents overfitting and provides the model with moderate flexibility. We demonstrate our model's capability on supervised action classification tasks.

II. Preliminaries

A. Gaussian Mixture Models

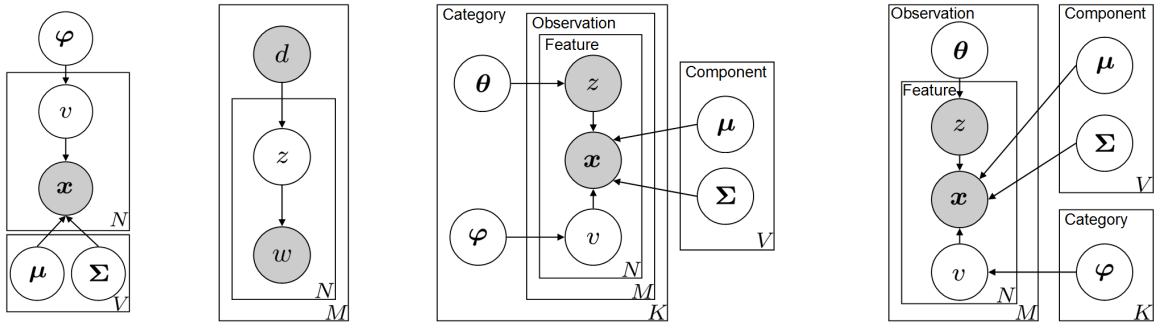
GMMs provide a richer class of density modeling than a single Gaussian distribution over continuous variables. The graphical representation of GMMs is depicted in Figure 1a. The latent variable v is the index of a Gaussian component that generates the observation $\boldsymbol{x} \in \mathbb{R}^{\|\boldsymbol{x}\|}$. GMMs encode the distribution of the *i.i.d.* observations $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$:

$$P(\boldsymbol{X}|\boldsymbol{\varphi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{j=1}^{N} \sum_{v=1}^{V} P(v|\boldsymbol{\varphi}) P(\boldsymbol{x}_j|\boldsymbol{\mu}_v, \boldsymbol{\Sigma}_v) \qquad (1)$$

where $P(v|\boldsymbol{\varphi})$ is the mixture coefficient, and $P(\boldsymbol{x}|\boldsymbol{\mu}_v, \boldsymbol{\Sigma}_v)$ is the multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_v, \boldsymbol{\Sigma}_v)$. We assume the elements in an observation vector are independent. Accordingly, the covariance matrix $\boldsymbol{\Sigma}$ becomes a diagonal matrix: $\boldsymbol{\Sigma}_v = \sigma_v^2 \boldsymbol{I}$, where $\boldsymbol{I}$ is the unit matrix.

B. Probabilistic Latent Semantic Analysis

The PLSA model [1] provides a probabilistic formulation for modeling topics over a document and a corpus. The graphical representation of the PLSA model is depicted in Figure 1b. Given the hidden variable z, each discrete word w is assumed to be independent of the document d which



(a) GMM (b) PLSA (c) Supervised HGMM (d) Supervised Gaussian PLSA

Figure 1: Plate representation of the graphical models discussed in this paper. The boxes are plates, representing replications.

Table I: Notations for our models

Notation	Meaning
$\boldsymbol{x}$	A feature vector with continuous elements
$\boldsymbol{X}$	The set of features from all observations
d	A dummy variable indexing an observation
z	The category assignment to $\boldsymbol{x}$
v	The Gaussian component assignment to $\boldsymbol{x}$
K	The number of topics
M	The number of observations in a category
N	The number of features in an observation
V	The number of Gaussian components
$\boldsymbol{\theta}$	Parameters of per-observation category distribution
$\boldsymbol{\varphi}$	Parameters of per-category feature distribution
$\boldsymbol{\mu}$	Mean of a Gaussian component
$\boldsymbol{\Sigma}$	Variance of a Gaussian component
$\boldsymbol{\Psi}$	Model parameters to learn: $\boldsymbol{\Psi} = \{\boldsymbol{\theta}, \boldsymbol{\varphi}\}$
$\boldsymbol{\Phi}$	All model parameters: $\boldsymbol{\Phi} = \{\boldsymbol{\Psi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$

contains it. The joint distribution of the observed variables (i.e., word and document variables) can be computed by:

$$P(d_i, w_j) = P(d_i) \sum_{z=1}^{K} P(z|d_i) P(w_j|z) \tag{2}$$

where $P(z|d_i)$ is the mixture weight, which is the probability that a topic z occurs in a document d, and $P(w_j|z)$ is the probability that a discrete word w occurs in a topic z. PLSAs treat each document as a mixture of topics, and treat each topic as a convex combination of discrete words.

III. Regularized Gaussian PLSAs

Our RG-PLSA model combines the advantages of PLSAs and GMMs, and minimizes or removes their disadvantages. We first introduce Hierarchical GMMs (HGMM) that are able to simultaneously model the distribution over features and categories. Then, we discuss our approaches to incorporate continuous features into PLSAs, and to also add regularization, which allows a tradeoff between overfitting and underfitting. Lastly, we show how to learn our model's parameters. For simplicity, we only focus on learning the mixture weights $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$, and we assume that the Gaussian component parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are learned beforehand and remain unchanged. All notations are listed in Table I.

A. Hierarchical GMMs

The HGMM model is graphically represented in Figure 1c. HGMMs explicitly model each continuous feature as a mixture of multivariate Gaussian components. Each category is also modeled as a mixture of the same Gaussian components. Thus, categories and features become dependent, and each category can be viewed as a mixture of the features. In addition, each category is also modeled as a multinomial distribution over all categories, in which the correct category assignment has the highest probability. Formally, for each category, the HGMM model represents the distribution:

$$P(\boldsymbol{X}|\boldsymbol{\Phi}) = \prod_{d=1}^{M} \prod_{i=1}^{N} \sum_{z=1}^{K} \sum_{v=1}^{V} P(z^{(d,i)}|\boldsymbol{\theta}) \\ P(v^{(d,i)}|\boldsymbol{\varphi}_z) P(\boldsymbol{x}^{(d,i)}|\boldsymbol{\mu}_{v,z}, \boldsymbol{\Sigma}_{v,z}) \tag{3}$$

where the superscript (d, i) denotes the ith feature in the dth observation, the subscript indicates which parameter is used, and $\boldsymbol{\Sigma}$ is assumed to be a diagonal matrix. It should be noted that, in HGMMs, the parameter $\boldsymbol{\theta}$ is fixed for each category, and does *not* depend on observations.

B. Gaussian PLSA

The Gaussian PLSA model replaces discrete words with continuous features that are modeled as a mixture of the multivariate Gaussian distributions, which is illustrated in Figure 1d. Each feature plate represents the distribution of the ith feature in an observation d along with its category assignment z and its Gaussian component assignment v. For the entire dataset, the joint distribution can be factorized as:

$$P(\boldsymbol{X}|\boldsymbol{\Phi}) = \prod_{d=1}^{M} \prod_{i=1}^{N} \sum_{z=1}^{K} \sum_{v=1}^{V} P(z^{(d,i)}|\boldsymbol{\theta}_d) \\ P(v^{(d,i)}|\boldsymbol{\varphi}_z) P(\boldsymbol{x}^{(d,i)}|\boldsymbol{\mu}_{v,z}, \boldsymbol{\Sigma}_{v,z}) \tag{4}$$

where $P(\boldsymbol{z}|\boldsymbol{\theta}_d)$ and $P(\boldsymbol{v}|\boldsymbol{\varphi}_z)$ are the multinomial distributions, and $P(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the multivariate Gaussian distribution with a diagonal covariance matrix. It should be noted that the only difference between Gaussian PLSAs and HGMMs is that the parameter $\boldsymbol{\theta}_d$ *does* depend on the observations in the Gaussian PLSA models.

C. Regulated Gaussian PLSAs

Our RG-PLSA model has the same graphical representation as the Gaussian PLSA model. However, in the parameter learning process, a regularization term is adopted to prevent overfitting. The EM algorithm is applied to iteratively learn model parameters, which is the most widely used frequentist parameter estimation in the latent variable graphical models. For each category, the RG-PLSA model's parameters can be learned by maximizing the regularized auxiliary function:

$$\begin{aligned} &Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^t,\boldsymbol{\Psi}_G^t) \triangleq P(\boldsymbol{X}|\boldsymbol{\Phi}) \\ &+\sum_{d=1}^{M}\sum_{i=1}^{N}\sum_{z=1}^{K}\sum_{v=1}^{V} P(z,v|\boldsymbol{\Psi}^t)\log\frac{P(\boldsymbol{x}^{(d,i)},z,v|\boldsymbol{\Phi})}{P(\boldsymbol{x}^{(d,i)},z,v|\boldsymbol{\Phi}^t)} \\ &+\lambda\sum_{d=1}^{M}\sum_{i=1}^{N}\sum_{z=1}^{K}\sum_{v=1}^{V} P(z,v|\boldsymbol{\Psi}_G^t)\log\frac{P(\boldsymbol{x}^{(d,i)},z,v|\boldsymbol{\Phi})}{P(\boldsymbol{x}^{(d,i)},z,v|\boldsymbol{\Phi}_G^t)} \\ &\propto \mathrm{E}_{P_R(\boldsymbol{z},\boldsymbol{v}|\boldsymbol{X},\boldsymbol{\Psi}^t,\boldsymbol{\Psi}_G^t)}[\log P(\boldsymbol{X},\boldsymbol{z},\boldsymbol{v}|\boldsymbol{\Phi})] \end{aligned} \tag{5}$$

where $\lambda \in [0,\infty)$ is the regularization factor that controls model complexity, and $\boldsymbol{\Psi}_G = \{\boldsymbol{\theta}_G, \boldsymbol{\varphi}\}$, $\boldsymbol{\Phi}_G = \{\boldsymbol{\Psi}_G, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$, i.e., only the per-observation category distribution is regulated. $P_R(z,v|\boldsymbol{X},\boldsymbol{\Psi}^t,\boldsymbol{\Psi}_G^t)$ is the regularized distribution over the latent variables, which can be computed by:

$$P_R(\boldsymbol{z},\boldsymbol{v}|\boldsymbol{X},\boldsymbol{\Psi}^t,\boldsymbol{\Psi}_G^t) = \frac{P(\boldsymbol{z},\boldsymbol{v}|\boldsymbol{X},\boldsymbol{\Psi}^t)+\lambda P(\boldsymbol{z},\boldsymbol{v}|\boldsymbol{X},\boldsymbol{\Psi}_G^t)}{1+\lambda} \tag{6}$$

The regularized distribution demonstrates the importance of the regularization factor λ: a smaller λ makes the RG-PLSA model behave more similarly to the Gaussian PLSA model, which allows for more model complexity. When $\lambda = 0$, the RG-PLSA model has the same form as the Gaussian PLSA model. Similarly, a larger λ emphasizes more on preventing overfitting, and HGMMs are a special instance of the RG-PLSA model, as $\lambda \to \infty$.

In the E-step, given the data and the current parameter values, the posterior distributions over the latent variables are computed:

$$w_{z,v}^{(d,i)} \triangleq P_R(z^{(d,i)}, v^{(d,i)}|\boldsymbol{x}^{(d,i)},\boldsymbol{\Psi}^t,\boldsymbol{\Psi}_G^t) \tag{7}$$

where we use $w_{z,v}^{(d,i)}$ as a simpler notation of this distribution.

In the M-step, new optimal parameter values are computed, given the re-estimated latent variables. Formally, the parameters are learned by:

$$\begin{aligned} \boldsymbol{\Psi}^{t+1} = \underset{\boldsymbol{\Psi}}{\mathrm{argmax}}(&Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^t,\boldsymbol{\Psi}_G^t) \\ &+\sum_{d=1}^{M}\delta_d(1-\sum_{z=1}^{K}\theta_{d,z})+\sum_{z=1}^{K}\delta_z(1-\sum_{v=1}^{V}\varphi_{z,v})) \end{aligned} \tag{8}$$

where the second and third terms are the Lagrange multipliers. Solving Equation (8) results in the parameter estimates:

$$\theta_{d,z}^{t+1} = \frac{\sum_{i=1}^{N}\sum_{v=1}^{V} w_{z,v}^{(d,i)}}{\sum_{z=1}^{K}\sum_{i=1}^{N}\sum_{v=1}^{V} w_{z,v}^{(d,i)}} \tag{9}$$

$$\varphi_{z,v}^{t+1} = \frac{\sum_{d=1}^{M}\sum_{i=1}^{N} w_{z,v}^{(d,i)}}{\sum_{v=1}^{V}\sum_{d=1}^{M}\sum_{i=1}^{N} w_{z,v}^{(d,i)}} \tag{10}$$

Finally, for each category, the regularized parameter estimate $\boldsymbol{\theta}_G$ is updated by:

$$\theta_{G_z}^{t+1} = \frac{\exp\left(\frac{1}{MN}\sum_{d=1}^{M}\sum_{i=1}^{N}\log\theta_{d,z}^{t+1}\right)}{\sum_{z'=1}^{K}\exp\left(\frac{1}{MN}\sum_{d=1}^{M}\sum_{i=1}^{N}\log\theta_{d,z'}^{t+1}\right)} \tag{11}$$

which is essentially the geometric mean of the observation-dependent $\theta_{d,z}$ in the same category. The log scale is applied to make the computation more manageable when $\theta_{d,z} \to 0$.

Given a new observation $\boldsymbol{Y} = \{\boldsymbol{y}_1,\ldots,\boldsymbol{y}_N\}$, the inference process selects the category that is most compatible with $\boldsymbol{Y}$. The first step is to estimate $\hat{\boldsymbol{\theta}}_d$ according to Equation (9), which depends on the observation. Then, with the estimated $\hat{\boldsymbol{\theta}}_d$, the RG-PLSA model chooses the category $C(\boldsymbol{Y})$ with the highest probability to generate the observation:

$$C(\boldsymbol{Y}) = \underset{c}{\mathrm{argmax}}\, P(\boldsymbol{Y}|\boldsymbol{\Phi}_c) \tag{12}$$

where $\boldsymbol{\Phi}_c = \{\hat{\boldsymbol{\theta}}_d, \boldsymbol{\varphi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}_c$ is the model parameter for the category c. It should be noted that the RG-PLSA model and the Gaussian PLSA model have the same inference process.

IV. Experiments

We evaluate our RG-PLSA model on the task of human action recognition, using two benchmark datasets: the Weizmann Action Dataset [3] and the KTH Action Dataset [4].

A. Methodology

We use the following pipeline to solve the task of human action recognition from a sequence of visual data:

1) Video preprocessing: Foreground-background segmentation can be used to obtain the regions of interests (ROIs) that contain humans. We directly use the ROIs provided by the datasets[1], since detecting ROIs is not our focus.

2) Feature extraction: Scale invariant feature transform (SIFT) [6] features are used to represent local human shape and appearance, which generate vectors with 128 elements. Then, principal component analysis is used to project each vector to a 10 dimensional feature vector.

3) Gaussian component learning: We fit the standard GMM model over the feature space. The number of Gaussian components is set to 600 for both datasets. The objective

[1]The bounding box data of the human subjects in the KTH Action Dataset are provided in [5], which are publicly available at: http://www.umiacs.umd.edu/~zhuolin/PrototypeTree/KTHBoundingBoxInfo.txt

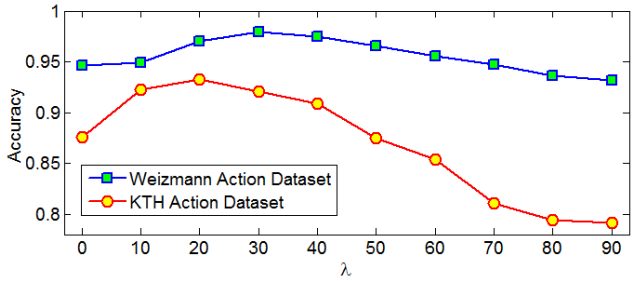


Figure 2: Average classification accuracy of the RG-PLSA model over the Weizmann and KTH Action Datasets.

Table II: Comparison of action classification accuracy over the Weizmann Action Dataset and the KTH Action Dataset

The Weizmann Dataset		The KTH Dataset	
Methods	Accuracy	Methods	Accuracy
RG-PLSAs	**0.9796**	**RG-PLSAs**	**0.9327**
Gaussian PLSAs	0.9465	Gaussian PLSAs	0.8760
HGMMs	0.9214	HGMMs	0.7878
Lin et al. [5]	1.0000	Lin et al. [5]	0.9577
Niebles et al. [7]	0.9000	Niebles et al. [7]	0.8150

of this step is to obtain the Gaussian components, which are fixed during the learning and inference processes.

4) Action classification: The RG-PLSA model is applied to categorize human actions. Evaluation is done using the leave-one-person-out cross-validation technique, in which videos of one human are used as the validation data, and videos from the remaining humans are used as the training data. This is repeated in such a way that videos from one human are used exactly once as the validation data.

B. Weizmann Action Dataset

Figure 2 illustrates the average classification accuracy of the proposed RG-PLSA model over the Weizmann dataset across different values of the regularization factor λ. When $\lambda = 30$, the RG-PLSA model achieves the best average accuracy of 97.96%, which outperforms the HGMM model and the Gaussian PLSA model, as shown in Table II. When $\lambda \to 0$, the RG-PLSA model tends to overfit the data, and the average accuracy deceases. With $\lambda = 0$, the RG-PLSA model obtains the same average accuracy as the Gaussian PLSA model. When λ becomes too large, the model tends to underfit the data, and the average accuracy tends to slowly decrease. We also provide the results from other approaches, as listed in Table II, to show that our method is comparable to the state of the art. We would like to emphasize that we are not making a direct comparison, because different approaches have variations in preprocessing, feature extraction, and experimental settings.

C. KTH Action Dataset

Figure 2 depicts our model's average classification accuracy over the KTH Action Dataset across different values of λ. With $\lambda = 20$, the best average classification accuracy of 93.27% is achieved. This result empirically demonstrates the benefit of applying the regularization term to the Gaussian PLSA model to improve classification accuracy. It should be noted that comparing to the results in the Weizmann dataset, the best result with the KTH dataset is achieved with a smaller λ, which allows for more model complexity and places more emphasis on the variations in the data. This occurs because the KTH dataset contains more variations and is more complicated than the Weizmann dataset. Our model shows similar behavior for both datasets with regard to λ in that very large λ values result in underfitting, and very small values lead to overfitting. We also compare other approaches over the KTH dataset in Table II, which shows that the RG-PLSA model achieves similar performance to other state-of-the-art methods.

V. Conclusion

We introduce the novel RG-PLSA model that combines the Gaussian PLSA model and the HGMM model to simultaneously prevent overfitting and provide moderate flexibility to model observations with continuous values. The proposed model employs a regularization term in the standard PLSA learning process to control model complexity. The RG-PLSA model's parameters are learned with the EM algorithm. We use two publicly available benchmark dataset to evaluate the effectiveness of our model on the human action recognition tasks. We achieve an accuracy of 97.96% for the Weizmann Action Dataset, and an accuracy of 93.27% for the KTH Action Dataset. These experimental results demonstrate that the proposed RG-PLSA model outperforms Gaussian PLSAs and HGMMs, and is comparable to the state of the art.

References

[1] T. Hofmann, "Probabilistic latent semantic analysis," in *Conf. Uncertainty in Artificial Intelligence*, 1999.

[2] S.-F. Wong, T.-K. Kim, and R. Cipolla, "Learning motion categories using both semantic and structural information," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2007.

[3] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," *Trans. Pattern Analysis and Machine Intel.*, vol. 29, no. 12, pp. 2247–2253, 2007.

[4] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: A local svm approach," in *Int'l Conf. Pattern Recognition*, 2004, pp. 32–36.

[5] Z. Lin, Z. Jiang, and L. S. Davis, "Recognizing actions by shape-motion prototype trees," in *IEEE Int'l Conf. Computer Vision*, 2009, pp. 444–451.

[6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int'l J. Computer Vision*, vol. 60, pp. 91–110, 2004.

[7] J. C. Niebles, H. Wang, and L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," *Int. J. Comput. Vision*, vol. 79, pp. 299–318, 2008.