

# Topic Modeling for Content Based Image Retrieval

Hemant Misra, Anuj K. Goyal and Joemon M. Jose

**Abstract** Latent Dirichlet allocation (LDA) topic model has taken a center stage in multimedia information retrieval, for example, LDA model was used by several participants in the recent TRECVID evaluation “Search” task. One of the common approaches while using LDA is to train the model on a set of test images and obtain their topic distribution. During retrieval, the likelihood of a query image is computed given the topic distribution of the test images, and the test images with the highest likelihood are returned as the most relevant images. In this paper we propose to project the unseen query images also in the topic space, and then estimate the similarity between a query image and the test images in the semantic topic space. The positive results obtained by the proposed method indicate that the semantic matching in topic space leads to a better performance than conventional likelihood based approach; there is an improvement of 25 % absolute in the number of relevant results extracted by the proposed LDA based system over the conventional likelihood based LDA system. Another not-so-obvious benefit of the proposed approach is a significant reduction in computational cost.

**Keywords** Multimedia information retrieval · Topic modeling · Latent Dirichlet allocation · Semantic information

---

H. Misra (✉)

Philips Research Asia, Philips Electronics India Limited, Bangalore, India  
e-mail: Hemant.Misra@philips.com

A. K. Goyal

Language Technologies Institute, Carnegie Mellon University, Pittsburgh, USA  
e-mail: Anuj@cs.cmu.edu

J. M. Jose

Department of Computing Science, University of Glasgow, Glasgow, UK  
e-mail: Joemon.Jose@glasgow.ac.uk

## 1 Introduction

The amount of personal multimedia data present on the internet is so huge that using traditional methods of annotating the data for retrieval is no more practical. There is a constant search for algorithms which can assist in unsupervised retrieval of multimedia data. The two important requirements of these algorithms are good retrieval performance and low computational cost.

Content based image retrieval (CBIR) involves extracting relevant digital images on the basis of their visual content [1–4]. TRECVID and Image-CLEF evaluations are typical meeting grounds for participants to showcase their image extraction algorithms and compare the performance of their algorithms on a common benchmark. In the recent TRECVID evaluations, quite a few participants have used latent Dirichlet allocation (LDA) [5, 6] topic model for the search task [7, 8]. Notably the top performing team in 2008 evaluations used LDA as one of the components in their system. Apart from TRECVID evaluations, LDA and probabilistic latent semantic analysis (PLSA) [9] have enjoyed prominence in many CBIR publications [1–3].

In most of the LDA based systems, the typical approach for retrieval is to compute the likelihood of a query image given the topic distribution of the test images, and return those test images as relevant which give the highest likelihood [1, 8]. However, as pointed out in [10] for the information retrieval task, likelihood based systems do not perform well on their own.

In this paper, we propose to project bag-of-words (BOW) representation of query images also in the LDA topic space; this has two major advantages: (1) we are able to capture the semantic information present in a query (relation among visual words of the query), (2) in the lower dimension LDA topic space, the cost associated with matching a query image with the test images is significantly lower as compared to the cost associated with a likelihood based approach. A brief description of the previous use of topic models for image retrieval is in Sect. 2.3.

The rest of the paper is organized as follows: In Sect. 2, we give a brief description of LDA model. The description of the proposed system and its main differences from the LDA based systems previously used for image retrieval tasks are provided in Sect. 3. Experimental setup of this paper is described in Sect. 4. In Sect. 5, we compare the performance of the two LDA based methods on TRECVID 2009 benchmark and analyze the results. Conclusions of this study are presented in Sect. 6.

## 2 Latent Dirichlet Allocation

The LDA model for the task of unsupervised topic detection was proposed in [5, 6]. The authors demonstrated the advantages of the LDA model vis-à-vis several other models, including multinomial mixture model [11] and probabilistic latent

semantic analysis (PLSA) [9]. Like most models of text, LDA uses the bag-of-words (BOW) representation of documents. The key assumptions of LDA are that *each document is represented by a topic distribution* and *each topic has an underlying word distribution*.

LDA is a generative model and specifies a probabilistic method for generating a new document. Assuming a fixed and known number of topics,  $T$ , for each topic  $t$ , a distribution  $\phi_t$  is drawn from a Dirichlet distribution of order  $V$ , where  $V$  is the vocabulary size. The first step in generating a document is to choose a topic distribution,  $\theta_{dt}$ ,  $t = 1 \dots T$ , for that document from a Dirichlet distribution of order  $T$ . Next, assuming that the document length is fixed, for each word occurrence in the document, a topic,  $z_i$ , is chosen from this topic distribution and a word is selected from  $\phi_{z_i}$ , the word distribution of the chosen topic. Given the topic distribution of the document, each word is drawn independently of every other word.

Therefore, the probability of  $w_i$ , the  $i$ th word token in document  $d$ , is:

$$P(w_i | \theta_d, \phi) = \sum_{t=1}^T P(z_i = t | \theta_d) P(w_i | z_i = t, \phi) = \sum_{t=1}^T \theta_{dt} \phi_{tw_i} \quad (1)$$

where  $P(z_i = t | \theta_d)$  is the probability that given the topic distribution  $\theta_d$ ,  $t$ th topic was chosen for the  $i$ th word token and  $P(w_i | z_i = t, \phi)$  is the probability of word  $w_i$  given topic  $t$ .

The likelihood of document  $d$  is a product of terms such as (1), and can be written as:

$$P(C_d | \theta_d, \phi) = \prod_{v=1}^V \left[ \sum_{t=1}^T (\theta_{dt} \phi_{tv}) \right]^{C_{dv}} \quad (2)$$

where  $C_{dv}$  is the count of word  $v$  in  $d$  and  $C_d$  is the word-frequency count in  $d$ .

## 2.1 LDA: Training

In the LDA training, the following two sets of parameters are estimated from a set of documents (train data): the topic distribution in each document  $d$  ( $\theta_{dt}$ ,  $d = 1 \dots D$ ,  $t = 1 \dots T$ ) and the word distribution in each topic ( $\phi_{tv}$ ,  $t = 1 \dots T$ ,  $v = 1 \dots V$ ). In this paper, Gibbs sampling [6] method is used to estimate these two distributions due to its better convergence and it being less sensitivity to initialization.  $\alpha$  and  $\beta$ , two hyper-parameters of the LDA model, define the non-informative Dirichlet priors on  $\theta$  and  $\phi$  respectively.

The training process for LDA model using Gibbs sampling is explained in [6]. For each word token in the training data, the probability of assigning the current word token to each topic is conditioned on the topic assigned to all other word tokens except the current word token. A topic is sampled from this conditional distribution and assigned to the current one. In every pass of Gibbs sampling, this

process of assigning a topic for all the word tokens in the training data constitutes one Gibbs sample. The initial Gibbs samples are discarded as they are not a reliable estimate of the posterior. For a particular Gibbs sample, the estimates for  $\theta$  and  $\phi$  are derived from the counts of hypothesized topic assignments as:

$$\phi_{tv} = \frac{J_{tv} + \beta}{\sum_{k=1}^V J_{tk} + V\beta} \text{ and } \theta_{dt} = \frac{K_{dt} + \alpha}{\sum_{k=1}^T K_{dk} + T\alpha}$$

where  $J_{tv}$  is the number of times word  $v$  is assigned to topic  $t$  and  $K_{dt}$  is the number of times topic  $t$  is assigned to some word token in document  $d$ .

## 2.2 LDA: Testing

In a typical information retrieval (IR) setting, where the main focus is on computing the similarity between a document  $d$  and a query  $d'$ , a natural similarity measure is given by  $P(C_{d'}|\theta_d, \phi)$ , computed according to (2) [12]. An alternative would be to compute the similarity through measures which are well suited for comparing distributions such as cosine distance, Bhattacharyya distance or KL divergence between  $\theta_d$  and  $\theta_{d'}$  (the topic distribution in  $d$  and  $d'$ ); this however requires to infer the latter quantity. As the topic distribution of a (new) document gives its representation along the latent semantic dimensions, computing this value is helpful for many applications such as language model adaptation [13] and text classification [14]. In [5], an approximate convexity based variational approach was proposed for inference. However, as pointed out in [15], the variational approach for inference has high bias and high computational cost.

In this paper, we use the expectation-maximization (EM) like iterative procedure suggested in [13, 14] for estimating topic distribution. The update rule is given by:

$$\theta_{dt} \leftarrow \frac{1}{l_d} \sum_{v=1}^V \frac{C_{dv} \theta_{dt} \phi_{tv}}{\sum_{t'=1}^T \theta_{dt'} \phi_{t'v}} \quad (3)$$

where  $l_d$  is document length, computed as the number of running words. It was shown in [14] that this update rule converges monotonically towards a local optimum of the likelihood, and the convergence is typically achieved in less than 10 iterations.

## 2.3 LDA: Application in CBIR

As mentioned previously, though the topic models were initially proposed for processing text [5, 6, 9] recently they have gained popularity in many other applications related to text processing [7, 10, 12–14, 16, 17] and image processing [1–3,

18–20]. PLSA based approaches [2, 3, 20] and LDA based approaches [1, 7, 8] have given good performance on very large databases. As was the case in text processing applications, LDA has typically yielded better performance than PLSA in image processing domain as well [1, 5].

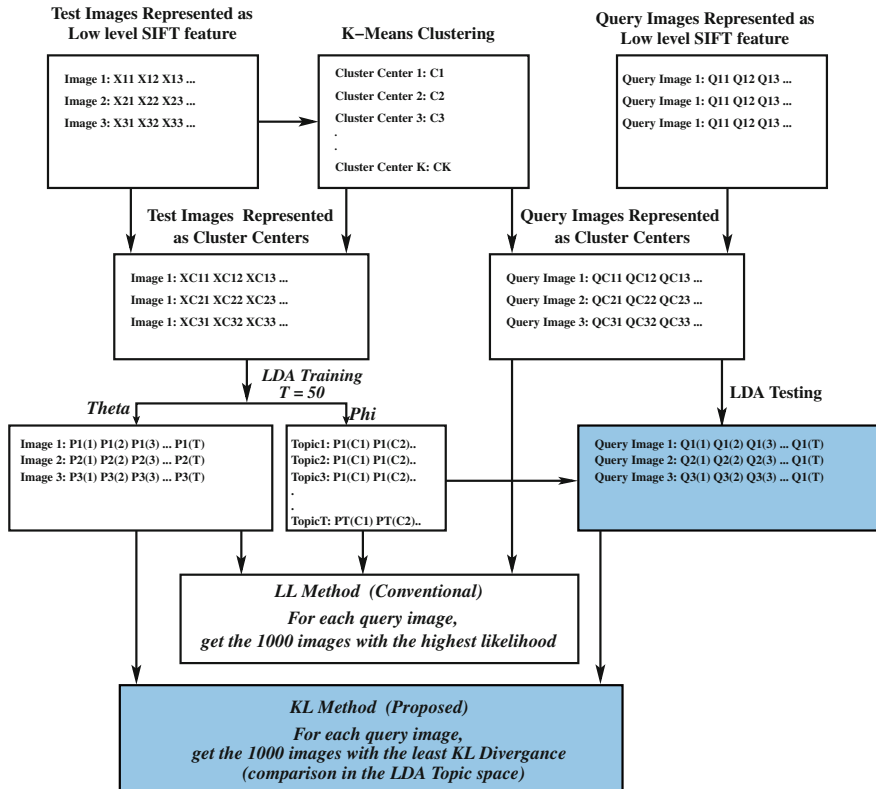
The work which is closest in nature to the idea presented in this paper is [1]. In Ref. [1], the authors did users studies to compare the performance of LDA based approach with PLSA based approach, and found LDA to perform better than PLSA. Also, they used the concept of relevance feedback to improve the results. The major differences between [1] and the present work are: (a) in [1], the authors used approximate convexity based variational approach [5] for computing the topic distribution of queries (unseen data). It has been pointed out in [15] that this approach is prone to bias and has high computational cost. In the present paper, we use the method presented in Sect. 2.2 for computing the topic distribution of queries. This method has been shown to have excellent performance on several other tasks [13, 14], and to the best of our knowledge is more accurate and much faster than any other methods proposed in the literature for LDA inference [15], (b) time complexity of the approach proposed in this paper is very low, thus making it possible to use this approach for online applications, (c) the results reported in this paper are on TRECVID 2009 benchmark in terms of standard measures such as mean average precision (MAP) and precision at 10 (P@10) making it possible to compare these results with other algorithms, whereas the results presented in [1] were on user evaluations.

### 3 System Description

Figure 1 shows the major components of the proposed and conventional LDA based methods; the most important difference is projecting the queries into the LDA topic space (LDA Testing block) and then doing the similarity match in the LDA topic space. Both these blocks are shaded in the figure.

#### 3.1 Low-Level Features

Our images are represented in terms of low-level local features obtained by Scale Invariant Feature Transform (SIFT) [21]. SIFT was the preferred choice in our case because we wanted a feature representation which is able to model regions of variable size in an image; also SIFT is invariant to scale, orientation and affine distortion and is partially immune to changes in illumination conditions.



**Fig. 1** Different components of the proposed system (**KL Method**: shaded blocks) and the conventional system (**LL Method**)

### 3.2 Clustering

LDA assumes that the documents to be modeled are represented in terms of a word vector; though the order of the words is not important, the words belong to a fixed vocabulary. Images represented in terms of features obtained by SIFT (henceforth we will call them SIFT features) cannot be used in an LDA model because the number of SIFT features is not limited. In order to limit the vocabulary size, we need to do some quantization of SIFT features and in our experiments we employed simple K-Means clustering to perform this desired quantization. We treated each SIFT feature as independent to obtain the  $K$  cluster centers. In the experiments reported in this paper, we have fixed  $K$ , the number of cluster centers (visual words), to be 10,000, which also becomes the size of our visual vocabulary.

We used the clustering code available in LIBPMK library [22] after making some necessary modifications. The time complexity of the K-Means algorithm increases with an increase in number of data points to be clustered; in order to keep

the computational cost under check and also to explore the dependence of the final performance on the amount of data used for clustering, we used 02, 05, 10 and 20 % of the *TRECVID 2008 relevant test data* for obtaining the 10,000 cluster centers. More details about this setup are provided in [Sect. 4](#).

The criterion for stopping the K-Means clustering in LIBPMK is the number of iterations. As the time complexity increases linearly with an increase in the number of iterations, in this paper we have explored the following four settings for the number of iterations: 10, 20, 40 and 80.

$L2 - norm$  was used to compute distances in the K-Means algorithm. The useful output of this clustering procedure is the  $K = 10,000$  cluster centers (visual words) which are used in the next stage of the proposed system.

### 3.2.1 Cluster Centers for 2009 Test Data

Once the visual words are obtained on a percentage of *TRECVID 2008 relevant test data*, representing the entire *TRECVID 2009 test data* in terms of these visual words is relatively simple: for each SIFT feature in each test image find the cluster center which is closest to it. We have used  $L2 - norm$  while estimating the distance between test SIFT features and cluster centers. At the end, each test image is represented in terms of cluster centers.

From a document perspective, each image is a document described by a word vector derived from a fixed vocabulary ( $K = 10,000$  cluster centers). This representation of images can be used for training the LDA model.

### 3.2.2 Cluster Centers for 2009 Query Data

Similar to the case of representing the *TRECVID 2009 test data* in terms of 10K visual words, one can represent each query image of the *TRECVID 2009 query data* in terms of visual words.

## 3.3 LDA Training on TRECVID 2009 Test Data

In this step, we train an LDA model on the entire TRECVID 2009 test collection represented in terms of  $K = 10,000$  cluster centers. *The hyper-parameters of the LDA model were  $\alpha = 1$ ,  $\beta = 0.1$  and number of topics,  $T = 50$ .* The procedure for training the LDA model was explained in [Sect. 2.1](#).

At the end of the training, the following two parameters of the LDA model are obtained: (1) each test image represented in terms of  $T$ -dimensional LDA topic distribution ( $\theta$ ), and (2) each LDA topic represented in terms of  $K$ -dimensional word distribution ( $\phi$ ). The LDA based approaches reported in the past followed all the steps up to this point. In the next section we briefly explain the difference

between the previous approaches and the approach proposed in this paper, highlighting the main advantages of the proposed approach.

### 3.4 Image Retrieval

As explained in [Sect. 2.2](#), in response to a query there are two possibilities to retrieve a set of images from a test collection.

1. **LL Method**: Compute the likelihood of the query given the images in the test collection and select those images as relevant that give the highest likelihood. In this case the likelihood is computed using (2) as follows (computation is typically done in log domain to avoid underflow that is why we work with log-likelihood (LL) instead of likelihood):  $P(C_q|\theta_d, \phi) = \prod_{v=1}^V [\sum_{t=1}^T (\theta_{dt} \phi_{tv})]^{C_{qv}}$ .
2. **KL Method**<sup>1</sup>: First estimate the topic distribution of the query image using the iterative procedure given in (3) as follows:  $\theta_{qt} \leftarrow \frac{1}{I_q} \sum_{v=1}^V \frac{C_{qv} \theta_{qt} \phi_{tv}}{\sum_{t'=1}^T \theta_{qt'} \phi_{t'v}}$ . This is an extremely fast procedure and convergence is typically reached in less than 10 iterations. Then symmetric KL divergence between  $\theta_q$ , the topic distribution of the query image, and  $\theta_d$ , the topic distribution of the test image  $d$  is computed. We select those test images as relevant that give the least KL divergence, where  $KL(\theta_q, \theta_d) = \sum_{t=1}^T [\theta_{dt} \log(\theta_{dt}/\theta_{qt}) + \theta_{qt} \log(\theta_{qt}/\theta_{dt})]$ .

By projecting the queries in the LDA topic space we are able to capture the semantics (relationship among words in a query) whereas this information is missing when each word in a query is treated independent of every other word.

The second, but not so obvious, advantage of the **KL Method** is the significant reduction in computational cost that can be realized by projecting the queries onto a lower dimensional LDA topic space and then doing the matching. The computational cost of the **LL Method** is dependent upon the query length which is significantly much higher than the number of LDA topics, specially when the images are represented by SIFT features. The average number of visual word tokens in a query were found to be 795 in the *TRECVID 2009 query data* collection. The high computational cost of **LL Method** was cited as its drawback in [1] as well. We will discuss more about the performance and the computational efficiency of the proposed **KL Method** in [Sect. 5](#).

---

<sup>1</sup> It was reported in [1] that cosine distance performs poorly as compared to KL divergence. In this paper we have considered symmetric KL divergence as the measure to estimate the similarity/distance between two images.



## 4 Experimental Setup

In TRECVID 2009 evaluations, approximately 280 hours of video data was provided as test set for the search task. The information about the shot boundaries was provided by NIST for the test set. In our experiments, we extracted one frame per shot. With this setup, our TRECVID 2009 evaluation test database for the search task consists of 97149 images. The query dataset had 471 images, either extracted from video shots or static images. In case of video, again we extracted only one frame per shot.

The number of search topics in TRECVID 2009 were 24. These topics can be considered as multimedia statement of information need. TRECVID results are typically reported as follows: given the search test collection, a topic (multimedia statement of information need of a user), and the common shot boundary reference for the search test collection, return a ranked list of at most  $N$  common reference shots from the test collection which best satisfy the user need. In TRECVID evaluations,  $N = 1,000$  for the standard search task whereas  $N = 10$  for the high-precision search task.

In our experiments, while performing K-Means clustering, we kept the number of cluster centers,  $K$ , fixed at 10,000. In the clustering algorithm, we studied the effect of the following two variables on the final retrieval performance:

- the number of iterations, and
- the amount of data from *TRECVID 2008 relevant test dataset* that was used for clustering.

We used the number of iterations as 10, 20, 40, and 80. The amount of data used for clustering was either 2, 5, 10 or 20 % of the relevant test dataset. For example, 2 % of the SIFT features from each image in the test collection were pooled together to create 2 % of the relevant test dataset for clustering. *The hyper parameters of the LDA model were kept fixed as  $\alpha = 1$ ,  $\beta = 0.1$  and  $T = 50$  in all the experiments.*

For each TRECVID topic a result list containing 1,000 shots was to be generated. When there were several query examples in a topic, each example generated a list of 1,000 most relevant shots. In such a case, each example was given equal importance and a voting was performed to generate the final list of 1,000 most relevant shots from these individual lists.

## 5 Results and Analysis

### 5.1 Retrieval Performance

In this section, we present the results of the two LDA based systems, one proposed in this paper and the other typically used in the literature. We present the performance

of the systems in terms of mean average precision (MAP), precision at R (R-prec), precision at 10 (P@10), precision at 1,000 (P@1,000) and total number of relevant results returned out of 10619 relevant results (Relevant).

Owing to the high computational cost associated with the **LL Method**, we were able to run only a few experiments for this system. In Table 1 we compare the performance of the two systems for two different number of iterations, 10 and 20.

It must be noted that the underlying LDA model for the **KL Method** and **LL Method** is exactly the same (the LDA model changes with the change in amount of data and number of iterations used for K-Means clustering).

Comparing the results of the two methods we observe that the proposed **KL Method** gives a better performance than **LL Method** across all the measures. This result is valid for two different training setups of the LDA model. Projecting the queries into the LDA topic space and then performing the similarity between the queries and the test documents in the LDA topic space not only gives a better performance, it also brings a significant reduction in cost complexity of the system. The time complexities of the two systems are described in the next section.

The results in Table 2 show the performance of **KL Method** for different *number of iterations used for K-Means clustering* when data used for clustering is 2 % of the total *TRECVID 2008 relevant test data*.

The results show a trend, though it is weak, that increasing the number of iterations of the K-means algorithm brings an improvement in the performance of the system by retrieving more relevant documents towards the end of the list (note that though P@1,000 improves, P@10 drops with an increase in number of iterations)

In Table 3, we present the performance of **KL Method** obtained by changing the *amount of data used for K-Means clustering*. The number of iterations were fixed at 10. Again we see a weak trend that increasing the amount of data used for K-Means clustering leads to a small improvement in performance. As in the previous case, though P@1,000 improves slightly, the P@10 drops, indicating that more relevant documents are added towards the end of the list. Also, for 20 % data size, performance drops slightly. The reason for this could be that as we increase

**Table 1** The performance of **KL Method** and **LL Method** for the same LDA model

Measure	<b>KL Method</b>		<b>LL Method</b>	
	Iterations		Iterations	
	10	20	10	20
MAP	0.0174	<b>0.0196</b>	0.0104	0.0110
R-prec	0.0405	<b>0.0433</b>	0.0340	0.0322
P@10	0.0792	<b>0.1083</b>	0.0375	0.0500
P@1,000	0.0424	<b>0.0456</b>	0.0360	0.0363
Relevant	1017	<b>1095</b>	863	872

The performance is shown for two different number of iterations, 10 and 20, whereas the data size is kept at 10 %. Performance in **bold** indicates that the improvement is statistically significant as compared to all the other systems

**Table 2** The performance of **KL Method** for different number of iterations in K-Means clustering

Measure	Iterations			
	10	20	40	80
MAP	0.0172	0.0183	0.0178	0.0177
R-prec	0.0388	0.0392	0.0412	0.0427
P@10	0.0917	0.0917	0.0708	0.0792
P@1,000	0.0408	0.0417	0.0425	0.0430
Relevant	979	1001	1019	1033

Data size = 2 % of the total TRECVID 2008 relevant test data

**Table 3** The performance of **KL Method** for different data sizes from TRECVID 2008 relevant test data used in K-Means clustering

Measure	% Data			
	2	5	10	20
MAP	0.0172	0.0178	0.0174	0.0166
R-prec	0.0388	0.0410	0.0405	0.0391
P@10	0.0917	0.0833	0.0792	0.0750
P@1,000	0.0408	0.0420	0.0424	0.0412
Relevant	979	1008	1017	990

Number of iterations is 10

the data size, we start selecting more data points which are non-relevant (like stop words) and these data points change the final cluster centers.

The encouraging result obtained from these experiments is that the performance of the *KL Method* is stable for different data sizes and different number of iterations which were used for clustering the data.

## 5.2 Time Complexity

In the proposed **KL Method**, the results are obtained in two steps: first we project each query image into the LDA topic space and then we compute the KL-divergence between the topic distribution of a query image and topic distribution of a test image. The test images are ranked based on their similarity to the query image in the topic space. Projecting a query image onto the LDA topic space is a very fast process and takes less than a second (298.7 seconds for 471 query images). LDA topic space has a much lower dimensionality than the test and query images represented in terms of cluster centers; as a consequence, the distance or similarity computation is very fast. Comparing this to the **LL Method** we find that though the test images are projected into the LDA topic space during LDA training, the query images are represented only in terms of cluster centers. Computing the

**Table 4** Time taken, in seconds, by **LL Method** and **KL Method**

<b>KL Method</b>		<b>LL Method</b>
Topic Estimation (of queries)	KL Divergence	LL Similarity
298.7 s	4544.16 s	76368 s

**KL Method** has two steps; time taken by the two steps is reported separately

likelihood of a query image with respect to all the test images requires that either (1) topic distribution of every test image ( $\theta_{dt}$ ) is multiplied with  $\phi_{tv}$  **only for the visual words which are present in the query image** to compute the likelihood of the test images with respect to the query, or (2) multiply  $\theta_{dt}$  of all the test images at once with  $\phi_{tv}$  and then **depending upon each query just sum the components which are relevant to that query**. (1) has less memory requirements but very high time requirements whereas (2) has high memory requirements ( $[NumberOfTestImages \times SizeOfVisualVocabulary]$ ) and moderate time requirements. Average time required by (2) and the **KL Method** are shown in Table 4. Memory requirement of (2) is too high to be processed on a simple machine and we used the grid facility provided by IRF, Vienna, to complete this task.

Comparing the results presented in Table 4 we observe that the proposed **KL Method** brings down the computational cost of the LDA approach by a factor of 15.7. The proposed LDA based approach not only gives an improvement of approximately 20 % over the conventional LDA based approach, it also reduces the time complexity by approximately 93.7 %.

It may also be noted that the LDA model was trained on TRECVID 2008 data whereas retrieval was performed on TRECVID 2009 data. Further improvement in the performance may be obtained if the model is trained and then used for retrieval on the same dataset. Moreover, it is possible to use more sophisticated clustering algorithms than K-Means to obtain a better visual vocabulary.

## 6 Conclusions

In this paper we proposed an LDA based system wherein the low-level SIFT features obtained from query images are first projected into the LDA topic space and then the matching between the query images and the test images is done in the LDA topic space. This is a departure from the conventional LDA based systems where typically the likelihood of the query images with respect to the test images is estimated to retrieve the most similar images. The proposed method not only leads to a significant improvement in the performance, it also reduces the computational cost associated with score estimation. In absolute terms, on TRECVID 2009 dataset, the number of relevant images retrieved by the proposed LDA system is approximately 20 % more than that obtained by the conventional

likelihood based LDA system. The reduction in computational cost while estimating the matching scores is more than 90 %. This result generalizes across all the training setups used in this study.

## References

1. Hörster E, Lienhart R, Slaney M (2007) Image retrieval on large-scale image databases. In ACM international conference on image and video retrieval, Amsterdam
2. Lienhart R, Slaney M (2007) PLSA on large-scale image databases. In IEEE international conference on acoustics, speech and signal processing, Honolulu, Hawaii
3. Monay F, Gatica-Perez D (2007) Modeling semantic aspects for cross-media image indexing. IEEE Transactions on Pattern Analysis and Machine Intelligence
4. Datta R, Joshi D, Li J, Wang JZ (2008) Image retrieval: ideas, influences, and trends of the new age. ACM computer surveys 40(2):1–60
5. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. Mach Learn Res 3:993–1022
6. Griffiths TL, Steyvers M (2004) Finding scientific topics. Proc Nat Acad Sci 101(supl 1):5228–5235
7. Cao J, Li J, Zhang Y, Tang S (2007) LDA-based retrieval framework for semantic news video retrieval. In IEEE international conference on semantic computing, Irvine, California, pp 155–160
8. Tang S, Li J-T, Li M, Xie C, Liu Y-Z, Tao K, Xu S-X (2008) TRECVID 2008 high-level feature extraction by MCG-ICT-CAS. In TRECVID 2008 Workshop. Gaithersburg, Maryland
9. Hofmann T (2001) Unsupervised learning by probabilistic latent semantic analysis. Mach Learn J 42(1):177–196
10. Wei X, Croft BW (2006) LDA-based document models for ad-hoc retrieval. In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, New York, p 178–185, ACM
11. Nigam K, McCallum AK, Thrun S, Mitchell TM (2000) Text classification from labeled and unlabeled documents using EM. Mach Learn 39(2/3):103–134
12. Buntine W, Löfström J, Perkiö J, Perttu S, Poroshin V, Silander T, Tirri H, Tuominen A, Tuulos V (2004) A scalable topic-based open source search engine. In Proceedings of the IEEE/WIC/ACM international conference on web intelligence, p 228–234, Beijing
13. Heidel A, an Chang H, shan Lee L (2007) Language model adaptation using latent Dirichlet allocation and an efficient topic inference algorithm. In proceedings of EuroSpeech, Antwerp, Belgium
14. Misra H, Cappé O, Yvon F (2008) Using LDA to detect semantically incoherent documents. In Proceedings of CoNLL, Manchester, pp 41–48
15. Yao L, Mimno D, McCallum A (2009) Efficient methods for topic model inference on streaming document collections. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, ACM, p 937–946
16. Xing D, Girolami M (2007) Employing latent Dirichlet allocation for fraud detection in telecommunications. Pattern recognition letters 28(13):1727–1734
17. Biró I, Siklósi D, Szabó J, Benczúr AA (2009) Linked latent Dirichlet allocation in web spam filtering. In Adversarial Information Retrieval on the Web, Madrid
18. Barnard K, Duygulu P, de Freitas N, Forsyth D, Blei D, Jordan MI (2003) Matching words and pictures. J Mach Learn Res 3:1107–1135
19. Blei DM, Jordan MI (2003) Modeling annotated data. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, New York, ACM, p 127–134

20. Bosch A, Zisserman A, Muñoz X (2006) Scene classification via pLSA. In European Conference on Computer Vision, p 517–530
21. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110
22. Lee JJ (2008) Libpmk: A pyramid match toolkit. Technical, Report MIT-CSAIL-TR-2008-17