People-LDA: Anchoring Topics to People using Face Recognition

Vidit Jain Erik Learned-Miller Andrew McCallum
University of Massachusetts Amherst
Amherst MA 01003
http://vis-www.cs.umass.edu/˜vidit/peopleLDA

Abstract

Topic models have recently emerged as powerful tools for modeling topical trends in documents. Often the resulting topics are broad and generic, associating large groups of people and issues that are loosely related. In many cases, it may be desirable to influence the direction in which topic models develop. In this paper, we explore the idea of centering topics around people. In particular, given a large corpus of images featuring collections of people and associated captions, it seems natural to extract topics specifically focussed on each person. What words are most associated with George Bush? Which with Condoleezza Rice? Since people play such an important role in life, it is natural to anchor one topic to each person.

In this paper, we present People-LDA, which uses the coherence of face images in news captions to guide the development of topics. In particular, we show how topics can be refined to be more closely related to a single person (like George Bush) rather than describing groups of people in a related area (like politics). To do this we introduce a new graphical model that tightly couples images and captions through a modern face recognizer. In addition to producing topics that are people specific (using images as a guiding force), the model also performs excellent soft clustering of face images, using the language model to boost performance. We present a variety of experiments comparing our method to recent developments in topic modeling and joint image-language modeling, showing that our model has lower perplexity for face identification than competing models and produces more refined topics.

1. Introduction

Topic models have recently emerged as powerful tools for modeling topical trends in documents. Often the resulting topics are broad and generic, associating large groups of people and issues that are loosely related. Typical topics that emerge from a set of newspaper articles might represent broad areas such as "sports", "politics", or "the Middle East". Of course, as large numbers of topics are extracted from a set of documents on the same narrow subject, topics will become more and more narrow, and "politics" may split into "the White House", "Capitol Hill", and "the Justice Department," or some comparable set of more focussed topics.

In many cases, it may be desirable to influence the direction in which topic models develop. In this paper, we explore the idea of centering topics around people. In particular, given a large corpus of images featuring collections of people and associated captions, it seems natural to extract topics specifically focussed on each person. What words are most associated with George Bush? Which with Condoleezza Rice? Since people play such an important role in life, it is natural to *anchor* one topic to each person. We use the term anchor to connote not only that a person should be a part of a topic, but that the topic should not drift too far from the topic defined by that person and their associations.

Below we present a new model, People-LDA, which uses the coherence of face images in news captions to guide the development of topics. In particular, we show how topics can be refined to be more closely related to a single person (like George Bush) rather than describing groups of people in a related area (like politics). To do this we introduce a new graphical model that tightly couples images and captions through a modern face recognizer.

Our model produces word topics that are people specific–it tends to eliminate secondary people or mixtures of people, focusing on a single person that matches a subset of face images. In addition, these people topics improve our ability to cluster faces over a method that uses only images.

In addition to producing topics that are people specific (using images as a guiding force), the model can also be used to cluster images by person, using the language model to boost performance. Our model has lower perplexity than competing models, meaning it assigns higher log probabilities, on average, to the correct name for a given face. We present a variety of experiments comparing our method to recent developments in topic modeling and joint image-language modeling.

Figure 1. An image from the "Faces in the wild" data set [3]. Associated caption: *President Bush, center, is flanked by the civilian U.S. administrator of Iraq L. Paul Bremer, right, and Secretary of Defense Donald H. Rumsfeld, left, as he makes remarks on Iraq, Wednesday, July 23, 2003, in the Rose Garden of the White House.*

1.1. Faces in the Wild

"Faces in the wild" [3] (see figure 1) is a data set that contains images and associated captions. The data set was extracted from news articles and contains images with a large amount of variation in pose, lighting, background and appearance. Some variation in appearance of faces of many people in this data set comes from motion (sports personalities) and make-up (Hollywood celebrities).

Berg et al. [3] first presented methods for clustering images from this data set, focusing particulary on the names of people written in each caption. They used a named-entity recognizer to extract names from the caption text, and then used analysis of the face images to choose one of the recognized names as the identity of each face in the corresponding image. While the accuracy of this method was impressive, there are a number of limitations to such an approach.

1. First, it relies heavily on the named-entity recognizer. These programs can be brittle, and it is very difficult to recover from missed names. These programs also cannot recognize that terms such as "the first lady" and "Laura Bush" may refer to the same person. If the name of a person does not appear in the caption, then the person cannot be identified.

2. Second, the method ignores important context and information provided by non-name text. Phrases like "Rose Garden" and "White House" (see figure 1) can provide critical context with which to identify difficult to recognize faces, even if the name of the pictured individual is not shown. By using such auxiliary information through carefully structured topic models, individuals can be identified even when their names do not directly appear in a given caption.

1.2. Additional Related Work

Latent Dirichlet allocation (LDA) [6] and its variants have been successful in modeling the generative process for text corpora. They have also been successfully adapted in several computer vision applications [8]. Barnard et al. [1] used a variant of LDA as a generative model for multimodal data (images and text). They showed that useful annotations can be obtained by modeling the joint distribution of images and associated text. In their work, the annotation words are broad categories of objects and background such as "sky", "grass", "building" and "people". They obtained promising results on general natural scene images. In this work, they used mixtures of Gaussian distributions to model these generic image-region categories.

In our task, we wish to develop significantly more precise and powerful models in order to identify specific faces. To this end, we adopt the hyper-feature classifier, which has been successfully used for the identification of faces and cars [9, 10]. In [9], Ferencz et al. showed that the hyper-feature model dramatically outperforms appearance models based upon mixture of Gaussian distributions. Instead of modeling the distribution in appearance of image patches/regions, Ferencz et al. [9] suggested modeling of the difference in appearance between a pair of image patches/regions. They modeled two distributions for difference in appearance between two images, one each for "same" and "different" people. The performance of this system was further improved by Jain et al. [10] by training the model discriminatively.

Because the hyper-feature model is a generative model of the *differences in appearance* of two face images, rather than a direct generative model of the appearance of a face, it is non-trivial to incorporate it into an LDA framework. A significant portion of our contribution represents the adjustment of the graphical model to accommodate the modeling of differences in appearance rather than appearance.

2. Hyper-feature based face identification

In an image, some patches are more useful than others in classifying the captured object into different classes like helicopter, car, face etc. These patches may not be very helpful in determining the identity within the detected class. Some type of image patches have similar appearance for all the objects in a given class, while other types of patches are specific to one (or few) objects in that class. We represent these patches by basic features like relative position, intensity values and edge energies in different directions. Using these features, we then learn to select the most informative patches and also estimate the variations in appearance of these patches for the same object. These basic features used to decide whether a patch would be useful in identifying a particular object or not, are called hyper-features. A hyper-feature based model was shown to be very effective for identifying objects like cars and faces [9].

For an object identification task, we are given a pair of images, (I_{left}, I_{right}). Let C be a binary random variable representing identification label as "same" or "different", and d be a continuous random variable representing the difference in appearance for a pair of image patches. We use h to represent the computed hyper-features for the left image, I_L, in the given image pair. The system discussed in [9] models $P(d|C =$"same"$, h)$ and $P(d|C =$"different"$, h)$ as gamma distributions that are trained independently of each other. Jain et al. [10] further improved this system by a more direct modeling of the desired identification criterion. They modeled the posterior probability of the label ("same"' or "different") given the image pair, $P(C =$"same"$|d, h)$, as a variant of logistic regression. For a face recognition task (on a data set consists of unconstrained images [3]), this system out-performed other recognizers [12]. In this work, we use a face identifier based on the approach used by Jain et al. [10].

3. People-LDA

For this discussion, we consider each document to be composed of an image $\mathbf{I}$, and a caption $\mathbf{w}$; a corpus represents a collection of D such documents. As shown in figure 2, each document is modeled as a mixture of people topics. In other words, for each document there may be more than one person appearing in the constituent image, and the associated caption text can be related to more than one individual.

People-LDA assumes the following generative process for each multi-modal document in a corpus D:

1. Choose a multinomial distribution θ over K people from a Dirichlet distribution. i.e. $\theta \sim Dir(\alpha)$, where α is a Dirichlet prior.
2. For n = 1 to N
   (a) Choose a person z_n from the chosen multinomial distribution in step 1. $z_n \sim Multinomial(\theta)$.
   (b) Choose a word w_n from a person specific distribution β_{z_n}.
3. For m = 1 to M
   (a) Choose a person z_{N+m} from the chosen multinomial distribution in step 1. $z_{N+m} \sim Multinomial(\theta)$.
   (b) For h = 1 to H
      i. Choose a patch I_h from the observed image $\mathbf{I}$ and compute its hyper-features.
      ii. Compute parameters Γ_h from a generalized linear model with parameter λ, i.e. $p(\Gamma_h|I_h, \lambda)$
      iii. Choose an appearance difference d_{mh} from a person-specific hyper-feature based distribution, $p(d_{mh}|z_{N+m}, \Gamma_h)$.

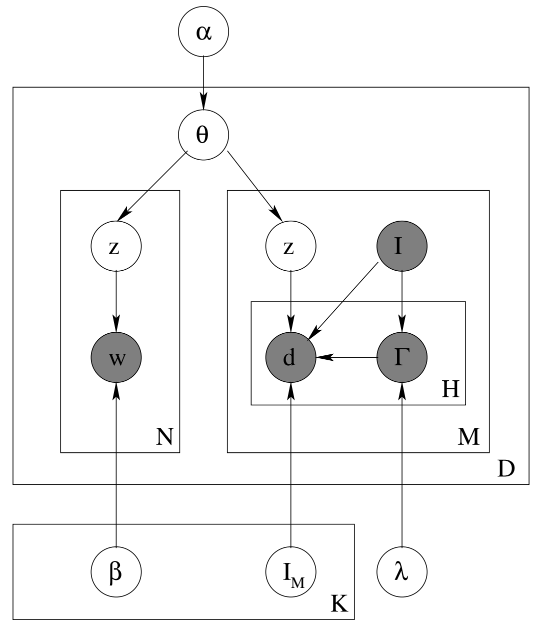


Figure 2. *People as topics*. The proposed topic model for modeling face images and associated caption text. Here, θ is a multinomial distribution, which specifies a mixture of people-topics z. For every document, we observe a number (M) of detected face images $\mathbf{I}$ and a caption, which is a collection of N words $\mathbf{w}$. For every face image, we sample H patches and compute the appearance difference d between I and a model-image (one of I_M). Γ represents the estimated parameters for face identifier as described in section 2. The overall parameters for the model are α, β, λ and a collection of K fixed model images, one for each person.

Given the parameters α, β and λ and an observed image $\mathbf{I}$, the joint distribution of a topic mixture θ, a set of $N+M$ topics $\mathbf{z}$, a set of N given words $\mathbf{w}$, M detected faces, image difference, $\mathbf{d}$, between model face image $\mathbf{I_M}$ is given by

$$p(\theta, \mathbf{z}, \mathbf{w}, \mathbf{d}|\alpha, \beta, \lambda, \mathbf{I}) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta) \cdot \prod_{m=1}^{M} p(z_{N+m}|\theta) \prod_{h=1}^{H} p(d_{mh}|z_{N+m}, \Gamma_h)p(\Gamma_h|\mathbf{I}, \lambda). \quad (1)$$

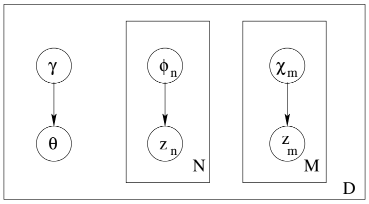


Figure 3. *Variational model for the proposed topic model.* λ, $\phi_{1:N}$, $\chi_{1:M}$ are the free parameters.

3.1. Variational inference

The exact probabilistic inference is intractable in our model as we cannot compute the posterior distribution (like LDA [6]). Here we use a fully factorized model, shown in figure 3, for the latent variables θ and $\mathbf{z}$:

$$q(\theta, \mathbf{z}) = q(\theta|\gamma) \prod_{n=1}^{N} q(z_n|\phi_n) \prod_{m=1}^{M} q(z_m|\chi_m) \quad (2)$$

with free parameters γ, ϕ and χ. γ is a K-dimensional Dirichlet parameter, ϕ_n and χ_m are N and M K-dimensional multinomial parameters, respectively. We then minimize the KL-divergence between the distribution $q(\theta, \mathbf{z})$ and the true posterior $p(\theta, \mathbf{z}|\mathbf{w}, \mathbf{I}, \alpha, \beta, \lambda, I_M)$. Taking derivatives of the KL-divergence with respect to the variational parameters, we obtain the following update equations:

$$\phi_{ni}^{t+1} = \beta_{iw_n} \exp(\Psi(\gamma_i^t)) \quad (3)$$

$$\chi_{mi}^{t+1} = p(d_m|z_{N+m} = i, \mathbf{I}, \lambda) \cdot \exp(\Psi(\gamma_i^t)) \quad (4)$$

$$\gamma_i^{t+1} = \alpha_i^{t+1} + \sum_{n=1}^{N} \phi_n^{t+1} + \sum_{m=1}^{M} \chi_m^{t+1}, \quad (5)$$

where $\Psi(.)$ is the digamma function. After every round of updates ϕ_n and χ_m are normalized so that they remain valid multinomial parameters.

3.2. Parameter estimation

For a given multi-modal corpus, we determine the maximum likelihood estimates of the model parameters using a modification of the variational EM procedure (see [2] for more details). The parameter estimation becomes complicated due to our choice of the form of distributions to represent the image component. To circumvent this, we train the face identifier (described in section 2) separately and assume the parameter λ to be fixed while estimating parameters α and β for our model. This corresponds to finding maximum likelihood parameters from a restricted set. Nevertheless, our experiments in section 5 demonstrate the effectiveness of this approach for parameter estimation. The E-step infers the variational parameters for each document given the current model parameters. The M-step computes maximum likelihood estimates of the model parameters using the variational distribution provided by the E-step. Our M-step updates are similar to that of Blei et al. [6].

4. Implementation details

In this section, we discuss some of the details of applying our proposed model (described in section 3) to obtain people-centered topics from a multi-modal data set. We consider a data set containing photographs, each having possibly more than one person appearing in it, and associated caption composed of one or two sentences of plain text. This is typical of photographs appearing in news articles.

4.1. Unsupervised selection of reference images

Our proposed model, People-LDA, requires at least one reference image per cluster. To obtain the reference images, we select images with only one face appearing in it and only one name present in the corresponding caption text. We use the Viola-Jones detector [13] for faces. A conditional random field based named entity recognizer [11] is used to extract names from the caption text. From this initial selection, we randomly choose one example per name as the corresponding reference image. Note that the named entity recognizer is used only for selecting reference images, and not for processing the caption while training or testing the model. Thus, our method is not particularly sensitive to the quality of the chosen named-entity recognizer.

4.2. Inferred distribution of topics for a document

For every document (image and caption, combined), our model infers a distribution over all the possible "people-topics". From the graphical model for People-LDA (figure 2), it is not obvious that the inferred mixture θ of people-topics will capture the co-occurrence relationship between detected faces and identified names in the caption text. However, the proper parameter values for the multinomial θ (when most of the probability mass is on one value) do indeed force a correspondence between names and faces. Since we are learning parameters of the model from the data itself, the inferred mixture θ for a document takes the desired form. A similar approach was used by Barnard et al. (MoM-LDA) [1] for annotating image with categorical words like "sky", "grass" and "water".

4.3. Annotation of faces in an image

For automatic annotation of faces in an image, we need conditional probabilities of words (names in our task) given the face appearance. A good model of joint probability of images and captions does not necessarily provide good estimates of conditional probabilities. MoM-LDA is a useful model for joint distribution of images and text, it may

Figure 4. *Reference image selection*. These images are selected automatically as discussed in section 4.1.

not be effective for annotation of faces with names. Blei et al. [4] suggested a model (Correspondence-LDA) which captures the correspondence between words and image regions, and thus models the conditional probability distribution of words given a region.

In People-LDA, the probability of a name given the topic is provided by a specialized face identifier, which separately optimizes the conditional probability distribution for classification. The computed probabilities are used in inferring the distribution over people-topic for a given face image (see equation 4 in section 3.1). For this reason, we do not need to specify an explicit correspondence between words and face images in our model.

4.4. "UNKNOWN" class

People-LDA annotates the faces in the given images with one of the selected names (section 4.1). In principle, it can not associate names for people whose reference images are not present. Thus, we need to filter out those documents (image and caption) from the given corpus where the chosen people are unlikely to appear in the image. Still, other people may appear along with the chosen people in a single image. To handle these cases, we use an additional "UNKNOWN" class as annotation for faces of people whose reference images are not selected.

5. Experiments

In our experiments, we used 10000 images and associated captions from the "Faces in the wild" data set [3]. Using the unsupervised selection of reference images (discussed in section 4.1), we obtain 1077 distinct names (of people). For our experiments, we randomly select 25 names in the middle frequency range (20-80 occurrences). These names can intuitively be categorized as related to sports; Pete Sampras, politics; Jacques Chirac and entertainment; Winona Ryder. The automatically selected reference images are shown in figure 4.

Our face identifier uses the ten most informative patches, chosen on the basis of the expected mutual information between the appearance difference and identification label. We trained the identification system on a set of 500 "same" and 500 "different" pair of images selected from the same "Faces in the wild" data set. Note that the training set does not contain images of the 25 people used in our experiments. This further demonstrates the merit of our approach as learning from one example, and suggests that our system should work for larger number of clusters as well. For text processing, we remove very frequent words ("stop-words" like "a", and "the") using a generic list, which is not specific to our document collection.

We compare People-LDA to approaches that use only the images (eigenfaces-Fisherfaces approach [14], hyper-feature model [10]) or only the captions (latent Dirichlet allocation [6]) and approaches that use both images and captions (Berg et al. [3] and Barnard et at. [1]).

5.1. People-Topics

We presented People-LDA as a model that guides topics to automatically emerge around people. In this section, we demonstrate this by comparing the image clusters that correspond to different people-topics and the topic specific word distributions for different approaches.

In particular, we compare the following three approaches (see figures 5 and 6): (a) *Image alone*: for each image, we use our face identifier to compute the probability of matching it with the reference images (one image per person), and assign this image to the cluster with maximum match probability. (b) *Text alone*: we first cluster the caption text using LDA [6]. For each caption, we assign the face images in the corresponding image to the most likely name under the inferred multinomial distribution of topics, and the learned topic-specific word distributions. (c) *People-LDA*: clustering obtained using our model.

Furthermore, in figure 1, we compare the ten most likely words under the distributions for different people-topics that are learned using LDA and People-LDA.

5.2. Quantitative Evaluation

To quantitatively evaluate the annotation quality of different models, we manually labeled the images used in our experiments and computed the perplexity of true label under different models. We also report the average class accuracy for classification (hard assignment).

As shown in table 2, joint modeling of images and text outperformed all the approaches that model images or text alone. Also note that the face recognizer used in our model [10] significantly outperformed the other face recognizer. We also implemented a naïve approach that randomly

(a) Random samples from four clusters obtained using face recognition [10] on images.

(b) The corresponding clusters obtained by People-LDA.

Figure 5. *Clustering using (a) image-only, and (b) People-LDA*. White squares are drawn manually on top of some of the images to highlight the number of distinct people in a cluster. The clusters are cleaned up significantly using our model and have fewer different people in them. *Other clusters are shown at the URL mentioned at the start of the paper*.

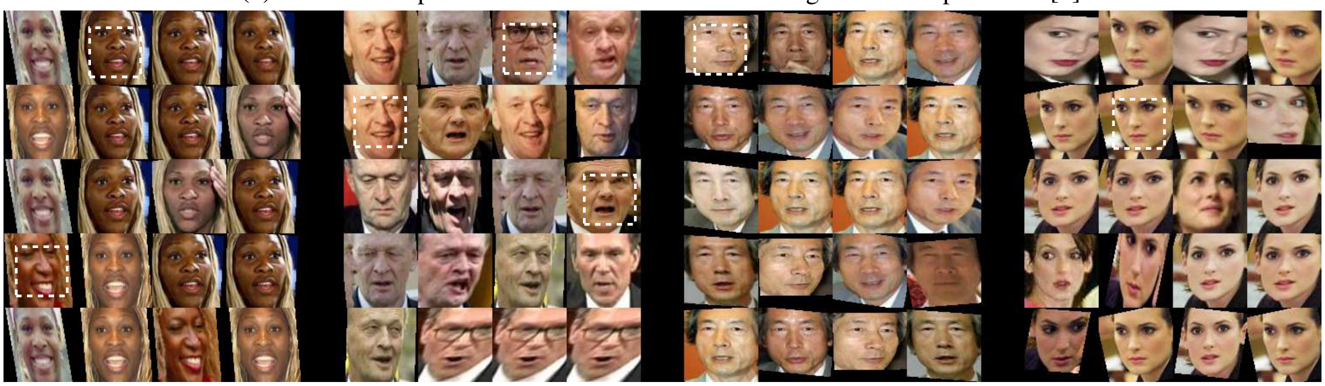
LDA				People-LDA			
schumacher	**chretien**	versace	**williams**	**schumacher**	**chretien**	**spears**	**williams**
chirac	**bush**	**chretien**	tennis	germany	**jean**	film	cup
koizumi	**jean**	**spears**	cup	cabinet	house	city	women
prix	street	poses	final	france	west	star	player
grand	cargo	**jean**	won	grand	ottawa	premiere	practice
michael	michigan	**britney**	**uribe**	**jean**	hill	poses	tennis
palace	facility	shows	returns	position	vote	**britney**	left
japan	suicide	women	development	announced	action	**watts**	number
jacques	fort	italian	tokyo	**michael**	question	mexico	montreal
french	detroit	final	princess	driver	government	week	week

Table 1. *Comparison of most likely words for people topics obtained by two models*. Each column corresponds to a topic learned by the model (LDA on caption text only or People-LDA). The name words are shown in bold face. These are four representative topics obtained using LDA. Topics obtained using People-LDA are more centered around one person compared to the topics for LDA. Moreover, the most likely name in a topic corresponds to the associated reference image. *Other topics are shown at the URL mentioned at the start of the paper*.

assigns one of the detected names in the caption to the faces detected in the corresponding image. This approach showed acceptable results as many captions have only one name detected in them.

People-LDA achieved the lowest label perplexity (lower values are better) among the compared methods. Both People-LDA and Barnard et al. [1] outperform the approach used by Berg et al. [3] since they model the probability distribution over all the possible names as compared to only the names detected in the caption only (as done by Berg

(a) Random samples from four clusters obtained using LDA on caption text [6].



(b) The corresponding clusters obtained by People-LDA.

Figure 6. *Clusters obtained using (a) text-only, and (b) People-LDA*. White squares are drawn manually on top of some of the images to highlight the number of distinct people in a cluster. The clusters are cleaned up significantly using our model and have fewer different people in them. *Other clusters are shown at the URL mentioned at the start of the paper*.

Model	Perplexity	% accuracy
Image Only		
Zhao et al. [14]	520.00 ± 24.17	22.02 ± 6.11
Hyper-features [10]	173.90 ± 3.96	44.86 ± 4.30
Text Only		
Random name from the caption	382.05 ± 23.11	31.40 ± 3.82
LDA on captions [6]	1219.60 ± 202.53	39.07 ± 2.44
Image and Text		
Barnard et al. [4]	68.23 ± 1.38	50.63 ± 4.01
Corr-LDA [4]	65.77 ± 2.13	52.50 ± 2.88
Berg et al. [3]	73.05 ± 9.36	68.93 ± 4.69
People-LDA	25.99 ± 4.50	58.56 ± 3.59

Table 2. *Quantitative evaluation*: In first column, we show the perplexity of the true label under different models (lower values are better). In the second column, the average class accuracies are shown. The error terms correspond to 10-fold cross-validation.

.

et al.). On the other hand, the method used by Berg et al. had the best average class accuracies among the compared methods. Their method draws advantage from the fact that many captions have a single name present in them (similar to our naïve approach). Furthermore, their approach fails to annotate a face if the corresponding name is not present in the caption (for example, see figure 7).

For a perfect labeling of all the faces in the data set, we still need to correct the misclassified faces. To do this, Berg et al. suggested the cost of correcting clustered data as a evaluation metric for different approaches. An alternative view of this cost is to consider only a few top matches and compute the recall (fraction of true labels present) of a system. In figure 8, our proposed model outperforms the other approaches.

6. Conclusion and Future Work

We proposed People-LDA as a model that guides semantic topics to develop around people. We achieved this by combining two successful models: hyper-feature based face identifier and latent Dirichlet allocation, in a novel way. As per our knowledge, no such combination has been proposed

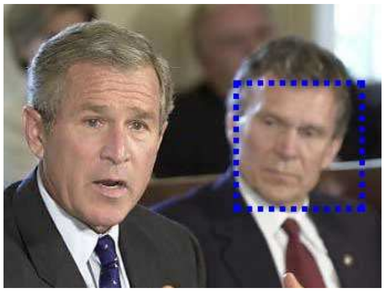


Figure 7. *A typical failure case for Berg et al. [3]*: Associated caption: *President George W. Bush (L) speaks to reporters at the conclusion of a bipartisan congressional meeting, September 4, 2002 at the White House. Bush asked Congress for nearly $1 billion to aid Israel and the Palestinians, fight the spread of AIDS and bolster security at U.S. airports.* Since the name "Tom Daschle" is not present in the caption, Berg et al. do not consider it as a possible label for the detected face in the given image.

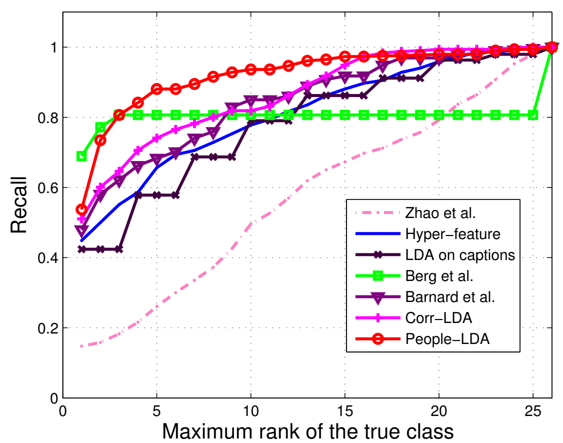


Figure 8. *Comparison of recall for different approaches when only top few matches are considered.* People-LDA outperforms the other approaches over most of the range. The approach used by Berg et al. [3] shows promising recall up to rank three but levels out as it does not consider names not present in the caption (none of the captions in our data set had more than three distinct names detected in them).

for joint modeling of images and text. We show excellent results of generation of people specific topics from a data set containing images and associated captions. Our model outperformed different modern approaches in soft clustering of face images.

There are several issues with LDA that affect the performance of our proposed model. First is the assumption that topics are uncorrelated. This causes the clustering results to be sensitive to the number of topics chosen, particularly for a large number of latent topics. Several richer models [5] have been proposed to overcome this weakness. Another issue with LDA arises if we have a highly skewed distribution of cluster frequencies. This causes the very frequent terms to appear in multiple clusters. To avoid this problem in our implementation, the most frequent terms (stop-words in the captions) were removed. Also, in our selected data set we are considering individuals who appeared with roughly the same frequently as others. Recently. Elkan [7] proposed a topic model that addresses this issue of frequency skewness. Our future work includes exploring such richer models for multi-modal documents.

Acknowledgements. We thank Marwan Mattar for useful discussions and suggestions. This work was supported in part by the Center for Intelligent Information Retrieval, the Central Intelligence Agency, the National Security Agency and National Science Foundation under CAREER award IIS-0546666 and grant IIS-0326249. Any opinions, findings and conclusions or recommendations expressed in this material are the authors' and do not necessarily reflect those of the sponsor.

References

[1] K. Barnard, P. Duygulu, D. A. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan. Matching words and pictures. *JMLR*, 3:1107–1135, 2003.

[2] M. J. Beal. Variational algorithms for approximate Bayesian inference. 2003.

[3] T. L. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Y. Teh, E. Learned-Miller, and D. A. Forsyth. Names and faces in the news. *CVPR*, 2004.

[4] D. Blei and M. Jordan. Modeling annotated data. In *SIGIR*, 2003.

[5] D. Blei and J. Lafferty. Correlated topic models. In *NIPS*, 2006.

[6] D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. *JMLR 3*, 2003.

[7] C. Elkan. Clustering documents with an exponential-family approximation of the Dirichlet compound multinomial distribution. In *ICML*, 2006.

[8] L. Fei-Fei and P. Perona. A Bayesian hierarchical model for learning natural scene categories. In *CVPR*, 2005.

[9] A. Ferencz, E. Learned-Miller, and J. Malik. Building a classification cascade for visual identification from one example. In *ICCV*, 2005.

[10] V. Jain, A. Ferencz, and E. Learned-Miller. Discriminative training of hyper-feature models for object identification. In *BMVC*, pages 357–366, 2006.

[11] A. McCallum and W. Li. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In *CoNLL*, 2003.

[12] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. *Pattern Recognition*, 33(11), 2000.

[13] P. Viola and M. J. Jones. Robust real-time face detection. *IJCV*, 57(2):137–154, May 2004.

[14] W. Zhao, R. Chellappa, and A. Krishnaswamy. Discriminant analysis of principal components for face recognition. In *International Conference on Automatic Face and Gesture Recognition*, 1998.