People-LDA using Face Recognition

Lijie Heng

12/11/2007

Outline

- Background
- Latent Dirichlet allocation
- People-LDA
- Experiments
- Conclusion

Background

- Modeling text corpora –Latent Dirichlet allocation (LDA)
- Newspaper articles (including captions + images)

  captions->LDA->topic

  images->face recognition->people
- Could we built a joint model on both image and text information?

Latent Dirichlet allocation

- In the text corpora, assume

  a word <- vocabulary{1,2,…V}

  a documents <- N words

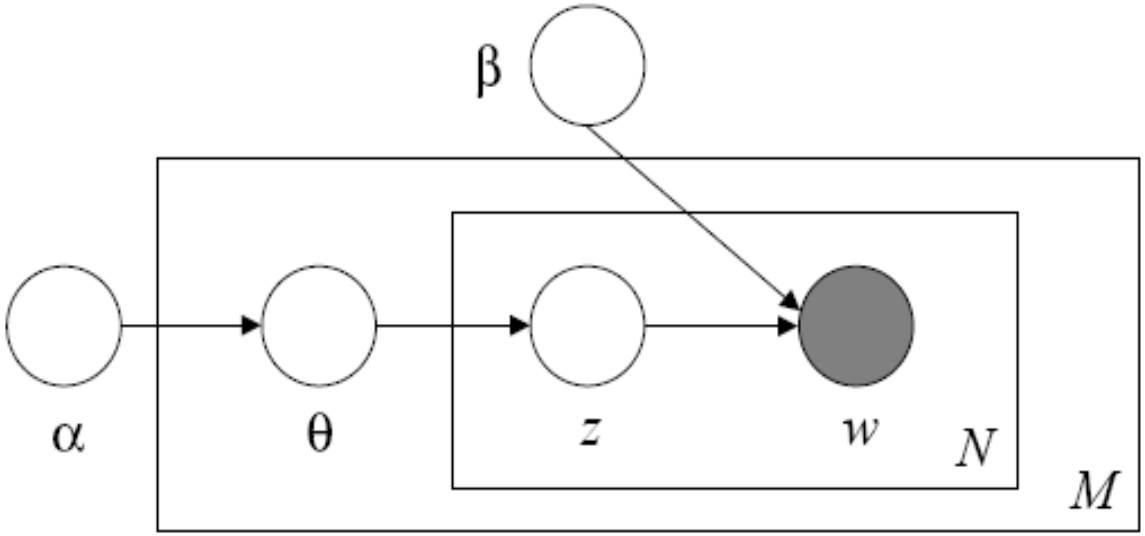
  a corpus <- M documents

Latent Dirichlet allocation-cont.

- To generate a document, we assume each document is generated from *K* topics and each topic is from *N* words from the vocabulary

1. Choose *N* ~ Poisson(x).
2. Choose θ ~ Dir(a).
3. For each of the *N* words *w_n*:
   (a) Choose a topic *z_n* ~ Multinomial(θ).
   (b) Choose a word *w_n* from *p*(*w_n* | *z_n; β*), a multinomial probability conditioned on the topic *z_n*.

Latent Dirichlet allocation-cont.



Latent Dirichlet allocation-cont.

Given the parameters α and β, the joint distribution of a topic mixture θ, a set of N topics $\mathbf{z}$, and a set of N words $\mathbf{w}$ is given by:

$$p(\theta, \mathbf{z}, \mathbf{w} \mid \alpha, \beta) = p(\theta \mid \alpha) \prod_{n=1}^{N} p(z_n \mid \theta) p(w_n \mid z_n, \beta),$$

$$p(\mathbf{w} \mid \alpha, \beta) = \int p(\theta \mid \alpha) \left(\prod_{n=1}^{N} \sum_{z_n} p(z_n \mid \theta) p(w_n \mid z_n, \beta) \right) d\theta.$$

$$p(D \mid \alpha, \beta) = \prod_{d=1}^{M} \int p(\theta_d \mid \alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn} \mid \theta_d) p(w_{dn} \mid z_{dn}, \beta) \right) d\theta_d.$$

People-LDA

- Take into account of image information in the documents
- Anchor each topic to a single person

  politics->George Bush, sports->Yao Ming

People-LDA cont.

- Assumptions

1. D documents in the corpus
2. K topics/people inside the corpus
3. Each document includes an image I and a caption W
4. Image I includes M faces, each faces contains H patches

People-LDA cont.

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}.

People-LDA cont.

3. For $m = 1$ to M

   (a) Choose a person z_{N+m} from the chosen multinomial distribution in step 1. $z_{N+m} \sim Multinomial(\theta)$.

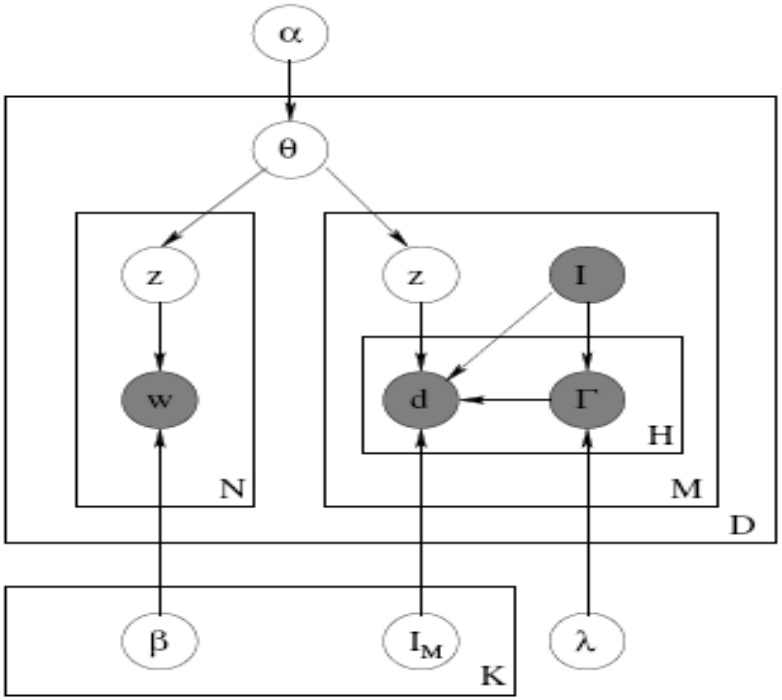
   (b) For $h = 1$ to H

      i. Choose a patch I_h from the observed image 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, $p(d_{mh} | z_{N+m}, \Gamma_h)$.

People-LDA cont.



$$p(\theta, \mathbf{z}, \mathbf{w}, \mathbf{d}|\alpha, \beta, \lambda, \mathbf{I}) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta)$$

$$\cdot \prod_{m=1}^{M} p(z_{N+m}|\theta) \prod_{h=1}^{H} p(d_{mh}|z_{N+m}, \Gamma_h)p(\Gamma_h|\mathbf{I}, \lambda). \quad (1)$$

Experiments

- Experiments:

1. 10000 documents from “Face in the wild”;
2. randomly select 25 names from 1077 distinct names showing in 10000 documents;
3. Obtain 25 reference faces(one image per person) as Reference Image
4. do image clustering

Experiments

- Image alone: using face identifier to clustering each image into one of the reference images
- Text alone: first cluster the caption text using LDA. then for each caption, assign the face images to their most likely names under the multinomial distribution of topics
- People-LDA

Experiments-*Clustering*

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

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

Experiments-*Clustering*

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

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

Experiments- C*lassification*

- Manually label the test images
- Compare the result image with the true label
- Report accuracy and perplexity(lower perplexity assigns higher the probability to correct images)

Experiments- C*lassification*

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.

Conclusion

- It’s a novel joint modeling of image and text.
- It has a better performance than other approaches.
- It can not associate names for people, whose reference images are not present.