Latent Dirichlet allocation

In natural language processing, **Latent Dirichlet Allocation (LDA)** is a generative model that allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar. For example, if observations are words collected into documents, it posits that each document is a mixture of a small number of topics and that each word's creation is attributable to one of the document's topics. LDA is an example of a topic model and was first presented as a graphical model for topic discovery by David Blei, Andrew Ng, and Michael Jordan in 2003.

Topics in LDA

In LDA, each document may be viewed as a mixture of various topics. This is similar to probabilistic latent semantic analysis (pLSA), except that in LDA the topic distribution is assumed to have a Dirichlet prior. In practice, this results in more reasonable mixtures of topics in a document. It has been noted, however, that the pLSA model is equivalent to the LDA model under a uniform Dirichlet prior distribution.

For example, an LDA model might have topics that can be classified as **CAT_related** and **DOG_related**. A topic has probabilities of generating various words, such as *milk*, *meow*, and *kitten*, which can be classified and interpreted by the viewer as "CAT_related". Naturally, the word *cat* itself will have high probability given this topic. The **DOG_related** topic likewise has probabilities of generating each word: *puppy*, *bark*, and *bone* might have high probability. Words without special relevance, such as *the* (see function word), will have roughly even probability between classes (or can be placed into a separate category). A topic is not strongly defined, neither semantically nor epistemologically. It is identified on the basis of supervised labeling and (manual) pruning on the basis of their likelihood of co-occurrence. A lexical word may occur in several topics with a different probability, however, with a different typical set of neighboring words in each topic.

Each document is assumed to be characterized by a particular set of topics. This is akin to the standard bag of words model assumption, and makes the individual words exchangeable.

Model

With plate notation, the dependencies among the many variables can be captured concisely. The boxes are "plates" representing replicates. The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document. M denotes the number of documents, N the number of words in a document. Thus:

Plate notation representing the LDA model.

α is the parameter of the Dirichlet prior on the per-document topic distributions,
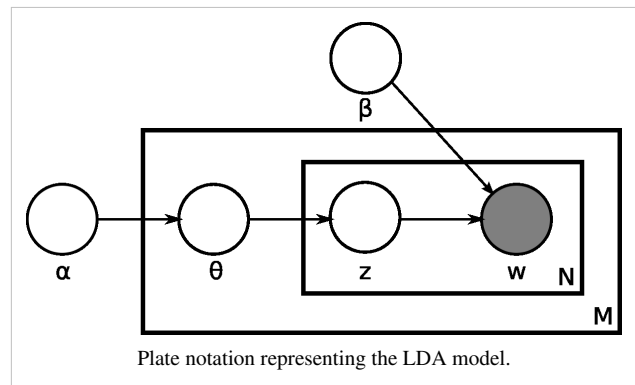
β is the parameter of the Dirichlet prior on the per-topic word distribution,

θ_i is the topic distribution for document i,

ϕ_k is the word distribution for topic k,

z_{ij} is the topic for the jth word in document i, and

w_{ij} is the specific word.

The w_{ij} are the only observable variables, and the other variables are latent variables. Mostly, the basic LDA model will be extended to a smoothed version to gain better results [citation needed]. The plate notation is shown on the right, where K denotes the number of topics considered in the model and:

Plate notation for smoothed LDA

> ϕ is a *K*V* (*V* is the dimension of the vocabulary) Markov matrix each row of which denotes the word distribution of a topic.
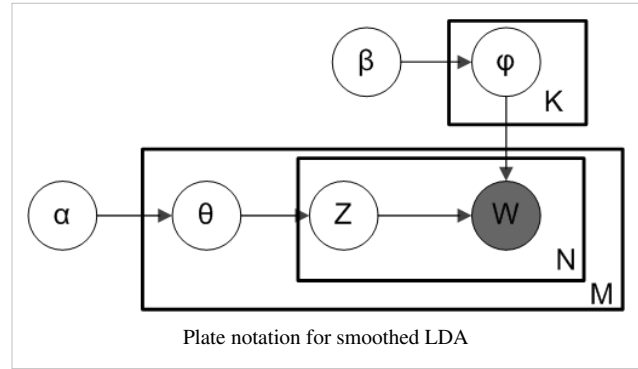
The generative process behind is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. LDA assumes the following generative process for a corpus D consisting of M documents each of length N_i:

1. Choose $\theta_i \sim \mathrm{Dir}(\alpha)$, where $i \in \{1, \ldots, M\}$ and $\mathrm{Dir}(\alpha)$ is the Dirichlet distribution for parameter α

2. Choose $\phi_k \sim \mathrm{Dir}(\beta)$, where $k \in \{1, \ldots, K\}$

3. For each of the word positions i, j, where $j \in \{1, \ldots, N_i\}$, and $i \in \{1, \ldots, M\}$

> (a) Choose a topic $z_{i,j} \sim \mathrm{Multinomial}(\theta_i)$.
> (b) Choose a word $w_{i,j} \sim \mathrm{Multinomial}(\phi_{z_{i,j}})$.

(Note that the Multinomial distribution here refers to the Multinomial with only one trial. It is formally equivalent to the categorical distribution.)

The lengths N_i are treated as independent of all the other data generating variables (q and z). The subscript is often dropped, as in the plate diagrams shown here.

Mathematical definition

A formal description of smoothed LDA is as follows:

Definition of variables in the model

Variable	Type	Meaning
K	integer	number of topics (e.g. 50)
V	integer	number of words in the vocabulary (e.g. 50,000 or 1,000,000)
M	integer	number of documents
$N_{d=1\ldots M}$	integer	number of words in document *d*
N	integer	total number of words in all documents; sum of all N_d values, i.e. $N = \sum_{d=1}^{M} N_d$
$\alpha_{k=1\ldots K}$	positive real	prior weight of topic *k* in a document; usually the same for all topics; normally a number less than 1, e.g. 0.1, to prefer sparse topic distributions, i.e. few topics per document
$\boldsymbol{\alpha}$	*K*-dimension vector of positive reals	collection of all α_k values, viewed as a single vector
$\beta_{w=1\ldots V}$	positive real	prior weight of word *w* in a topic; usually the same for all words; normally a number much less than 1, e.g. 0.001, to strongly prefer sparse word distributions, i.e. few words per topic
$\boldsymbol{\beta}$	*V*-dimension vector of positive reals	collection of all β_w values, viewed as a single vector
$\phi_{k=1\ldots K, w=1\ldots V}$	probability (real number between 0 and 1)	probability of word *w* occurring in topic *k*

$\phi_{k=1\ldots K}$	V-dimension vector of probabilities, which must sum to 1	distribution of words in topic k
$\theta_{d=1\ldots M, k=1\ldots K}$	probability (real number between 0 and 1)	probability of topic k occurring in document d for a given word
$\boldsymbol{\theta}_{d=1\ldots M}$	K-dimension vector of probabilities, which must sum to 1	distribution of topics in document d
$z_{d=1\ldots M, w=1\ldots N_d}$	integer between 1 and K	identity of topic of word w in document d
$\mathbf{Z}$	N-dimension vector of integers between 1 and K	identity of topic of all words in all documents
$w_{d=1\ldots M, w=1\ldots N_d}$	integer between 1 and V	identity of word w in document d
$\mathbf{W}$	N-dimension vector of integers between 1 and V	identity of all words in all documents

We can then mathematically describe the random variables as follows:

$$
\begin{aligned}
\phi_{k=1\ldots K} &\sim \text{Dirichlet}_V(\boldsymbol{\beta}) \\
\boldsymbol{\theta}_{d=1\ldots M} &\sim \text{Dirichlet}_K(\boldsymbol{\alpha}) \\
z_{d=1\ldots M, w=1\ldots N_d} &\sim \text{Categorical}_K(\boldsymbol{\theta}_d) \\
w_{d=1\ldots M, w=1\ldots N_d} &\sim \text{Categorical}_V(\boldsymbol{\phi}_{z_{dw}})
\end{aligned}
$$

Inference

Learning the various distributions (the set of topics, their associated word probabilities, the topic of each word, and the particular topic mixture of each document) is a problem of Bayesian inference. The original paper used a variational Bayes approximation of the posterior distribution; alternative inference techniques use Gibbs sampling and expectation propagation.

Following is the derivation of the equations for collapsed Gibbs sampling, which means φ s and θ s will be integrated out. For simplicity, in this derivation the documents are all assumed to have the same length N. The derivation is equally valid if the document lengths vary.

According to the model, the total probability of the model is:

$$
P(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\varphi}; \alpha, \beta) = \prod_{i=1}^{K} P(\varphi_i; \beta) \prod_{j=1}^{M} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t} | \theta_j) P(W_{j,t} | \varphi_{Z_{j,t}}),
$$

where the bold-font variables denote the vector version of the variables. First of all, $\boldsymbol{\varphi}$ and $\boldsymbol{\theta}$ need to be integrated out.

$$
P(\boldsymbol{Z}, \boldsymbol{W}; \alpha, \beta) = \int_{\boldsymbol{\theta}} \int_{\boldsymbol{\varphi}} P(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\varphi}; \alpha, \beta)\, d\boldsymbol{\varphi}\, d\boldsymbol{\theta}
$$

$$
= \int_{\boldsymbol{\varphi}} \prod_{i=1}^{K} P(\varphi_i; \beta) \prod_{j=1}^{M} \prod_{t=1}^{N} P(W_{j,t} | \varphi_{Z_{j,t}})\, d\boldsymbol{\varphi} \int_{\boldsymbol{\theta}} \prod_{j=1}^{M} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t} | \theta_j)\, d\boldsymbol{\theta}.
$$

All the θ s are independent to each other and the same to all the φ s. So we can treat each θ and each φ separately. We now focus only on the θ part.

$$
\int_{\boldsymbol{\theta}} \prod_{j=1}^{M} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t} | \theta_j)\, d\boldsymbol{\theta} = \prod_{j=1}^{M} \int_{\theta_j} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t} | \theta_j)\, d\theta_j.
$$

We can further focus on only one θ as the following:

$$\int_{\theta_j} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t}|\theta_j)\, d\theta_j.$$

Actually, it is the hidden part of the model for the jth document. Now we replace the probabilities in the above equation by the true distribution expression to write out the explicit equation.

$$\int_{\theta_j} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t}|\theta_j)\, d\theta_j = \int_{\theta_j} \frac{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} \theta_{j,i}^{\alpha_i - 1} \prod_{t=1}^{N} P(Z_{j,t}|\theta_j)\, d\theta_j.$$

Let $n_{j,r}^i$ be the number of word tokens in the jth document with the same word symbol (the rth word in the vocabulary) assigned to the ith topic. So, $n_{j,r}^i$ is three dimensional. If any of the three dimensions is not limited to a specific value, we use a parenthesized point $(\cdot)$ to denote. For example, $n_{j,(\cdot)}^i$ denotes the number of word tokens in the ith document assigned to the ith topic. Thus, the right most part of the above equation can be rewritten as:

$$\prod_{t=1}^{N} P(Z_{j,t}|\theta_j) = \prod_{i=1}^{K} \theta_{j,i}^{n_{j,(\cdot)}^i}.$$

So the θ_j integration formula can be changed to:

$$\int_{\theta_j} \frac{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} \theta_{j,i}^{\alpha_i - 1} \prod_{i=1}^{K} \theta_{j,i}^{n_{j,(\cdot)}^i}\, d\theta_j$$

$$= \int_{\theta_j} \frac{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} \theta_{j,i}^{n_{j,(\cdot)}^i + \alpha_i - 1}\, d\theta_j.$$

Clearly, the equation inside the integration has the same form as the Dirichlet distribution. According to the Dirichlet distribution,

$$\int_{\theta_j} \frac{\Gamma\left(\sum_{i=1}^{K} n_{j,(\cdot)}^i + \alpha_i\right)}{\prod_{i=1}^{K} \Gamma(n_{j,(\cdot)}^i + \alpha_i)} \prod_{i=1}^{K} \theta_{j,i}^{n_{j,(\cdot)}^i + \alpha_i - 1}\, d\theta_j = 1.$$

Thus,

$$\int_{\theta_j} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t}|\theta_j)\, d\theta_j = \int_{\theta_j} \frac{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} \theta_{j,i}^{n_{j,(\cdot)}^i + \alpha_i - 1}\, d\theta_j$$

$$= \frac{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \frac{\prod_{i=1}^{K} \Gamma(n_{j,(\cdot)}^i + \alpha_i)}{\Gamma\left(\sum_{i=1}^{K} n_{j,(\cdot)}^i + \alpha_i\right)} \int_{\theta_j} \frac{\Gamma\left(\sum_{i=1}^{K} n_{j,(\cdot)}^i + \alpha_i\right)}{\prod_{i=1}^{K} \Gamma(n_{j,(\cdot)}^i + \alpha_i)} \prod_{i=1}^{K} \theta_{j,i}^{n_{j,(\cdot)}^i + \alpha_i - 1}\, d\theta_j$$

$$= \frac{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \frac{\prod_{i=1}^{K} \Gamma(n_{j,(\cdot)}^i + \alpha_i)}{\Gamma\left(\sum_{i=1}^{K} n_{j,(\cdot)}^i + \alpha_i\right)}.$$

Now we turn our attentions to the φ part. Actually, the derivation of the φ part is very similar to the θ part. Here we only list the steps of the derivation:

$$\int_{\varphi} \prod_{i=1}^{K} P(\varphi_i; \beta) \prod_{j=1}^{M} \prod_{t=1}^{N} P(W_{j,t}|\varphi_{Z_{j,t}}) \, d\varphi$$

$$= \prod_{i=1}^{K} \int_{\varphi_i} P(\varphi_i; \beta) \prod_{j=1}^{M} \prod_{t=1}^{N} P(W_{j,t}|\varphi_{Z_{j,t}}) \, d\varphi_i$$

$$= \prod_{i=1}^{K} \int_{\varphi_i} \frac{\Gamma\left(\sum_{r=1}^{V} \beta_r\right)}{\prod_{r=1}^{V} \Gamma(\beta_r)} \prod_{r=1}^{V} \varphi_{i,r}^{\beta_r - 1} \prod_{r=1}^{V} \varphi_{i,r}^{n_{(\cdot),r}^{i}} \, d\varphi_i$$

$$= \prod_{i=1}^{K} \int_{\varphi_i} \frac{\Gamma\left(\sum_{r=1}^{V} \beta_r\right)}{\prod_{r=1}^{V} \Gamma(\beta_r)} \prod_{r=1}^{V} \varphi_{i,r}^{n_{(\cdot),r}^{i} + \beta_r - 1} \, d\varphi_i$$

$$= \prod_{i=1}^{K} \frac{\Gamma\left(\sum_{r=1}^{V} \beta_r\right)}{\prod_{r=1}^{V} \Gamma(\beta_r)} \frac{\prod_{r=1}^{V} \Gamma(n_{(\cdot),r}^{i} + \beta_r)}{\Gamma\left(\sum_{r=1}^{V} n_{(\cdot),r}^{i} + \beta_r\right)}.$$

For clarity, here we write down the final equation with both ϕ and θ integrated out:

$$P(\boldsymbol{Z}, \boldsymbol{W}; \alpha, \beta)$$

$$= \prod_{j=1}^{M} \frac{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \frac{\prod_{i=1}^{K} \Gamma(n_{j,(\cdot)}^{i} + \alpha_i)}{\Gamma\left(\sum_{i=1}^{K} n_{j,(\cdot)}^{i} + \alpha_i\right)} \times \prod_{i=1}^{K} \frac{\Gamma\left(\sum_{r=1}^{V} \beta_r\right)}{\prod_{r=1}^{V} \Gamma(\beta_r)} \frac{\prod_{r=1}^{V} \Gamma(n_{(\cdot),r}^{i} + \beta_r)}{\Gamma\left(\sum_{r=1}^{V} n_{(\cdot),r}^{i} + \beta_r\right)}.$$

The goal of Gibbs Sampling here is to approximate the distribution of $P(\boldsymbol{Z}|\boldsymbol{W}; \alpha, \beta)$. Since $P(\boldsymbol{W}; \alpha, \beta)$ is invariable for any of Z, Gibbs Sampling equations can be derived from $P(\boldsymbol{Z}, \boldsymbol{W}; \alpha, \beta)$ directly. The key point is to derive the following conditional probability:

$$P(Z_{(m,n)}|\boldsymbol{Z}_{-(m,n)}, \boldsymbol{W}; \alpha, \beta) = \frac{P(Z_{(m,n)}, \boldsymbol{Z}_{-(m,n)}, \boldsymbol{W}; \alpha, \beta)}{P(\boldsymbol{Z}_{-(m,n)}, \boldsymbol{W}; \alpha, \beta)},$$

where $Z_{(m,n)}$ denotes the Z hidden variable of the nth word token in the mth document. And further we assume that the word symbol of it is the vth word in the vocabulary. $\boldsymbol{Z}_{-(m,n)}$ denotes all the Z s but $Z_{(m,n)}$. Note that Gibbs Sampling needs only to sample a value for $Z_{(m,n)}$, according to the above probability, we do not need the exact value of $P(Z_{m,n}|\boldsymbol{Z}_{-(m,n)}, \boldsymbol{W}; \alpha, \beta)$ but the ratios among the probabilities that $Z_{(m,n)}$ can take value. So, the above equation can be simplified as:

$$P(Z_{(m,n)} = k|\boldsymbol{Z}_{-(m,n)}, \boldsymbol{W}; \alpha, \beta)$$

$$\propto P(Z_{(m,n)} = k, \boldsymbol{Z}_{-(m,n)}, \boldsymbol{W}; \alpha, \beta)$$

$$= \left(\frac{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}{\prod_{i=1}^{K} \Gamma(\alpha_i)}\right)^{M} \prod_{j \neq m} \frac{\prod_{i=1}^{K} \Gamma(n_{j,(\cdot)}^{i} + \alpha_i)}{\Gamma\left(\sum_{i=1}^{K} n_{j,(\cdot)}^{i} + \alpha_i\right)}$$

$$\times \left(\frac{\Gamma\left(\sum_{r=1}^{V} \beta_r\right)}{\prod_{r=1}^{V} \Gamma(\beta_r)}\right)^{K} \prod_{i=1}^{K} \prod_{r \neq v} \Gamma(n_{(\cdot),r}^{i} + \beta_r)$$

$$\times \frac{\prod_{i=1}^{K} \Gamma(n_{m,(\cdot)}^{i} + \alpha_i)}{\Gamma\left(\sum_{i=1}^{K} n_{m,(\cdot)}^{i} + \alpha_i\right)} \prod_{i=1}^{K} \frac{\Gamma(n_{(\cdot),v}^{i} + \beta_v)}{\Gamma\left(\sum_{r=1}^{V} n_{(\cdot),r}^{i} + \beta_r\right)}$$

$$\propto \frac{\prod_{i=1}^{K} \Gamma(n_{m,(\cdot)}^{i} + \alpha_i)}{\Gamma\left(\sum_{i=1}^{K} n_{m,(\cdot)}^{i} + \alpha_i\right)} \prod_{i=1}^{K} \frac{\Gamma(n_{(\cdot),v}^{i} + \beta_v)}{\Gamma\left(\sum_{r=1}^{V} n_{(\cdot),r}^{i} + \beta_r\right)}.$$

Finally, let $n_{j,r}^{i,-(m,n)}$ be the same meaning as $n_{j,r}^{i}$ but with the $Z_{(m,n)}$ excluded. The above equation can be further simplified by treating terms not dependent on k as constants:

$$\propto \frac{\prod_{i \neq k} \Gamma(n_{m,(\cdot)}^{i,-(m,n)} + \alpha_i)}{\Gamma\left(\left(\sum_{i=1}^{K} n_{m,(\cdot)}^{i,-(m,n)} + \alpha_i\right) + 1\right)} \prod_{i \neq k} \frac{\Gamma(n_{(\cdot),v}^{i,-(m,n)} + \beta_v)}{\Gamma\left(\sum_{r=1}^{V} n_{(\cdot),r}^{i,-(m,n)} + \beta_r\right)}$$

$$\times \Gamma(n_{m,(\cdot)}^{k,-(m,n)} + \alpha_k + 1) \frac{\Gamma(n_{(\cdot),v}^{k,-(m,n)} + \beta_v + 1)}{\Gamma\left(\left(\sum_{r=1}^{V} n_{(\cdot),r}^{k,-(m,n)} + \beta_r\right) + 1\right)}$$

$$\propto \frac{\Gamma(n_{m,(\cdot)}^{k,-(m,n)} + \alpha_k + 1)}{\Gamma\left(\left(\sum_{i=1}^{K} n_{m,(\cdot)}^{i,-(m,n)} + \alpha_i\right) + 1\right)} \frac{\Gamma(n_{(\cdot),v}^{k,-(m,n)} + \beta_v + 1)}{\Gamma\left(\left(\sum_{r=1}^{V} n_{(\cdot),r}^{k,-(m,n)} + \beta_r\right) + 1\right)}$$

$$= \frac{\Gamma(n_{m,(\cdot)}^{k,-(m,n)} + \alpha_k)\left(n_{m,(\cdot)}^{k,-(m,n)} + \alpha_k\right)}{\Gamma\left(\sum_{i=1}^{K} n_{m,(\cdot)}^{i,-(m,n)} + \alpha_i\right)\left(\sum_{i=1}^{K} n_{m,(\cdot)}^{i,-(m,n)} + \alpha_i\right)} \frac{\Gamma(n_{(\cdot),v}^{k,-(m,n)} + \beta_v)\left(n_{(\cdot),v}^{k,-(m,n)} + \beta_v\right)}{\Gamma\left(\sum_{r=1}^{V} n_{(\cdot),r}^{k,-(m,n)} + \beta_r\right)\left(\sum_{r=1}^{V} n_{(\cdot),r}^{k,-(m,n)} + \beta_r\right)}$$

$$\propto \frac{\left(n_{m,(\cdot)}^{k,-(m,n)} + \alpha_k\right)}{\left(\sum_{i=1}^{K} n_{m,(\cdot)}^{i,-(m,n)} + \alpha_i\right)} \frac{\left(n_{(\cdot),v}^{k,-(m,n)} + \beta_v\right)}{\left(\sum_{r=1}^{V} n_{(\cdot),r}^{k,-(m,n)} + \beta_r\right)}$$

$$\propto \left(n_{m,(\cdot)}^{k,-(m,n)} + \alpha_k\right) \frac{\left(n_{(\cdot),v}^{k,-(m,n)} + \beta_v\right)}{\left(\sum_{r=1}^{V} n_{(\cdot),r}^{k,-(m,n)} + \beta_r\right)}.$$

Note that the same formula is derived in the article on the Dirichlet-multinomial distribution, as part of a more general discussion of integrating Dirichlet distribution priors out of a Bayesian network.

Applications, extensions and similar techniques

Topic modeling is a classic problem in information retrieval. Related models and techniques are, among others, latent semantic indexing, independent component analysis, probabilistic latent semantic indexing, non-negative matrix factorization, and Gamma-Poisson_distribution.

The LDA model is highly modular and can therefore be easily extended. The main field of interest is modeling relations between topics. This is achieved by using another distribution on the simplex instead of the Dirichlet. The Correlated Topic Model follows this approach, inducing a correlation structure between topics by using the logistic normal distribution instead of the Dirichlet. Another extension is the hierarchical LDA (hLDA), where topics are joined together in a hierarchy by using the nested Chinese restaurant process. LDA can also be extended to a corpus in which a document includes two types of information (e.g., words and names), as in the LDA-dual model. Supervised versions of LDA include L-LDA, which has been found to perform better than SVM under certain conditions. Nonparametric extensions of LDA include the Hierarchical Dirichlet process mixture model, which allows the number of topics to be unbounded and learnt from data and the Nested Chinese Restaurant Process which allows topics to be arranged in a hierarchy whose structure is learnt from data.

As noted earlier, PLSA is similar to LDA. The LDA model is essentially the Bayesian version of PLSA model. The Bayesian formulation tends to perform better on small datasets because Bayesian methods can avoid overfitting the data. In a very large dataset, the results are probably the same. One difference is that PLSA uses a variable d to represent a document in the training set. So in PLSA, when presented with a document the model hasn't seen before, we fix $\Pr(w \mid z)$ --the probability of words under topics—to be that learned from the training set and use the same EM algorithm to infer $\Pr(z \mid d)$ --the topic distribution under d. Blei argues that this step is cheating because you are essentially refitting the model to the new data.

Variations on LDA have been used to automatically put natural images into categories, such as "bedroom" or "forest", by treating an image as a document, and small patches of the image as words; one of the variations is called Spatial Latent Dirichlet Allocation.

Notes

External links

- Extremely useful lecture for understanding LDA LDA and Topic Modelling Video Lecture by David Blei (http://videolectures.net/mlss09uk_blei_tm/)
- D. Mimno's LDA Bibliography (http://mimno.infosci.cornell.edu/topics.html) An exhaustive list of LDA-related resources (incl. papers and some implementations)
- Gensim Python+NumPy implementation of LDA for input larger than the available RAM.
- topicmodels (http://cran.r-project.org/web/packages/topicmodels/index.html) and lda (http://cran.r-project.org/web/packages/lda/index.html) are two R packages for LDA analysis.
- "Text Mining with R" including LDA methods (http://www.r-bloggers.com/RUG/2010/10/285/), video presentation to the October 2011 meeting of the Los Angeles R users group
- MALLET (http://mallet.cs.umass.edu/index.php) Open source Java-based package from the University of Massachusetts-Amherst for topic modeling with LDA, also has an independently developed GUI, the Topic Modeling Tool (http://code.google.com/p/topic-modeling-tool/)
- LDA in Mahout (https://cwiki.apache.org/confluence/display/MAHOUT/Latent+Dirichlet+Allocation) implementation of LDA using MapReduce on the Hadoop platform
- Perl implementation of LDA (http://www.people.fas.harvard.edu/~ptoulis/code/LDA_Perl.zip) A non-optimized implementation of the LDA model in Perl, with documentation
- Latent Dirichlet Allocation (LDA) Tutorial for the Infer.NET Machine Computing Framework (http://research.microsoft.com/en-us/um/cambridge/projects/infernet/docs/Latent Dirichlet Allocation.aspx) Microsoft Research C# Machine Learning Framework

Article Sources and Contributors

Latent Dirichlet allocation *Source*: http://en.wikipedia.org/w/index.php?oldid=583215337 *Contributors*: 3mta3, Aaron McDaid, AlanUS, Arcenciel, Artod, Benwing, Bkkbrad, Borice, Chiccodoro, Chris, Chzz, Comtebenoit, Danyaljj, DarikCR, Darklilac, Dennis.rippinger, Domthedude001, Drevicko, Drpickem, Euphoria83, Ezubaric, Facopad, Falsifian, Foma84, Foobarbazqux, Giftlite, Headbomb, Hornandsoccer, Jonsafari, Jrockley, Lauyukpui, Liontooth, Maghnus, Mcld, Melcombe, Michael Fourman, Michael Hardy, Qwertyus, Qwfp, Rama, Rdds, Rjwilmsi, Robinanil, Ronz, SPiNoZA, Shae, Slxu.public, Timohonkela, Took, Visviva, Yi.wang.2005, Ysaimanoj, Zeno Gantner, 91 anonymous edits

Image Sources, Licenses and Contributors

Image:Latent Dirichlet allocation.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Latent_Dirichlet_allocation.svg *License*: Creative Commons Attribution-Share Alike *Contributors*: Bkkbrad

Image:Smoothed LDA.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Smoothed_LDA.png *License*: Creative Commons Attribution-Sharealike 3.0 *Contributors*: Slxu.public

License