

Applying LDA in contextual image retrieval

ReDCAD participation at ImageCLEF Flickr Photo Retrieval 2012

Hatem Awadi, Mouna Torjmen Khemakhem, and Maher Ben Jemaa

Research unit on Development and Control of Distributed Applications (ReDCAD),
Department of Computer Science and Applied Mathematics,
National School of Engineers of Sfax, University of Sfax
{awadi.hatem@gmail.com, torjmen.mouna@redcad.org, maher.benjemmaa@enis.rnu.tn}
<http://www.redcad.org>

Abstract. This paper describes our participation in photo Flickr retrieval task at the ImageCLEF 2012 Campaign. Our aim is to evaluate the performance of topic models, such as Latent Dirichlet Allocation (LDA), in image retrieval based on the textual information surrounding the images. To do this, we propose to extract topics from Flickr user tags¹ using the LDA topic model. Then, we use the Jensen-Shannon Divergence measure to compute the similarity between queries and user tags representing images.

Keywords: text-based image retrieval, Latent Dirichlet Allocation, Jensen-Shannon Divergence

1 Introduction

Many works in the image retrieval literature have shown that, in Web case, textual retrieval is more efficient than content retrieval [8][9][14].

The common method of searching images by context is to use directly the text surrounding the images by applying the well known tf-idf scheme [11] which evaluates how important is a word in a document. While this approach reduces the document into a set of words that are discriminative for documents in the collection, it provides a relatively small amount of reduction in description length and does not capture inter- or intradocument statistical structure [2].

To resolve those problems, latent dimension can be used to reduce the term-document matrix to a much lower dimension subspace that captures most of the variance in the corpus. The main idea of this technique consists in modeling documents as a distribution of topics where each topic is a distribution of words.

In this paper, we choose to use Latent Dirichlet Allocation (LDA) [2] to model image topics. The first step is to extract topics from user tags representing images in the given Flickr collection, and then estimate topic distribution of

¹ User tags are a kind of metadata describing the images and allowing them to be found by searching or browsing

the query by inferring the query in the existing topics distributions. Finally, the Jensen-Shannon Divergence measure [7] is used to compute similarity between queries and user tags representing images.

Topic models are widely used in textual information processing and have shown their interest in many tasks. Recently, this technique was used in image representation and processing.

In the image retrieval domain, LDA is mainly used in visual level. Hoster et. al. [4] represented an image as a bag of visual words and then applied LDA to extract visual topics. Many similarity measures are tested where the Jensen-Shannon Divergence measure [7] performs the best. Greif et. al. [5] have also used a Correlated Topic Model (CTM [1]). However, this model did not perform over previous approaches.

Elango et. al. [3] used LDA topic model for image clustering. Another application of LDA is in automatic image annotation [10][13].

In this work, we are interested in applying LDA to the textual information related to the images. Resulted topics are then used to find images similar to a textual user query.

Our paper is organized as follows. Section 2 presents a review of LDA topic model and similarity measure that we use in image retrieval. In section 3, we present experimental results on photo Flickr retrieval task and conclude the paper in section 4.

2 LDA for image retrieval

The main idea behind the use of topic model in our work is that the image is probably an illustration of the overall subject (topic) in the document. User tags are likely to be motivating feature to represent the image since they normally describe the image content. For this, we use user tags to extract textual topics of the images. Figure 1 shows an example of a set of topics extracted from Flickr user tags.

ocean	space	art	phone	sunset
beach	stars	painting	mobile	evening
sea	chandra	museum	cell	sky
coast	galaxy	gallery	camera	dusk
pacific	smithsonian	artist	sony	clouds
water	institution	modern	cellphone	silhouette
waves	telescope	contemporary	ericsson	landscape
rocks	star	media	nokia	sun
surf	universe	mixed	telephone	twilight
sand	ray	collage	blackberry	sundown

Fig. 1. Top 10 words of 5 topics extracted from the Flickr user tags

2.1 Latent Dirichlet Allocation

In a large collection, the main problem is that many documents are about the same idea. Topic models are used to connect documents that share similar patterns (meaning) by discovering patterns of words.

The idea behind LDA is to model documents as a distribution of topics where each topic defines a distribution over words. Specifically, we assume that K topics are associated with a collection, and that each document defines a distribution over (hidden) topics. The posterior probability of these latent variables determines a hidden decomposition of the collection into topics.

We have D documents using a vocabulary of V word types. Each document contains M word tokens. We assume K topics. Each document has a K -dimensional multinomial θ over topics with a common Dirichlet prior $Dir(\alpha)$. Each topic has a V -dimensional multinomial ϕ over words with a common symmetric Dirichlet prior $Dir(\beta)$.

Figure 2 shows the various components of this model.

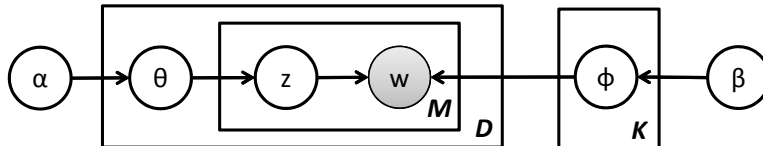


Fig. 2. A graphical model representation of the Latent Dirichlet Allocation (LDA)

The generative process of LDA is described as follow:

- (1) For each topic,
 - (a) Draw a distribution over words $\phi \sim Dir(\beta)$
- (2) For each document,
 - (a) Determine topic distribution $\theta_d \sim Dir(\alpha)$
 - (b) For each word,
 - (i) Generate topic $z \sim Mult(\theta)$
 - (ii) Generate word $w \sim Mult(\phi)$.

2.2 Similarity measure

After running LDA on a corpus, it is possible to use its output to compare documents to each other. In our case, each tag has a distribution over topics. Many works use the KL-divergence [6] to measure the distance between topics and therefore distance between documents as follows:

$$D_{KL}(P||Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)}. \quad (1)$$

where P and Q are two probability distributions over topics of two documents p and q .

But the problem is that the KL-divergence is not symmetric i.e. $D_{KL}(P||Q) \neq D_{KL}(Q||P)$. An example of symmetric divergence measure named Jensen-Shannon divergence [7] derived from KL divergence is widely used. To compare two distributions P and Q using Jensen-Shannon divergence, equation 2 is applied.

$$JSD(P||Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M). \quad (2)$$

where $M = \frac{1}{2}(P + Q)$.

By comparing documents in this way, two documents that share few words may be considered similar if they are about the same topic.

Equation 2 computes the distance between two distributions. To compute the similarity between a distribution Q of a query q and a distribution T of a tag t in our work, we use the following formula:

$$Similarity(Q, T) = 1 - JSD(Q||T). \quad (3)$$

Images related to the top similar tags are returned as a result to the query.

3 Experimental results

For the photo Flickr retrieval task, we use a subset of the MIRFLICKR² collection composed of 200 000 images. There are a number of 42 textual queries that are used to perform LDA-based image retrieval.

3.1 Parameter settings

Concerning the number of topics K , it can not be perfectly fixed because it depends on many factors, essentially the collection size (the number of documents). More the size of the collection increases, more the number of topics. So a large dataset needs a large K . In our experiment we fixed this number to 1000 since we have a large collection. We conserve the standard setting of the other parameters : $\alpha = 50/K$, $\beta = 0.01$.

3.2 Results

In this section, we present the results of our single official run of the LDA model. Table 1 shows obtained result [12].

According to the obtained results, our method does not perform very well compared to the best run in ImageCLEF2012 competition. A possible explication of this result is that the user query is generally composed of a few words. Consequently, we do not knew about its topic.

² <http://mirflickr.liacs.nl>

Table 1. Impact of LDA model in image retrieval

Run	MAP	AP@10	AP@20	AP@100
1341138375297_LDA01tx	0.0129	0.0003	0.0042	0.0475
Best run in ImageCLEF2012	0.0933	0.0187	0.0338	0.1715

4 Conclusion

This work studies the impact of textual topics in image retrieval. We have applied the LDA topic model to the user tags representing images. Results show that this approach does not perform very well. In future works, we plan to improve results by using the query expansion technique to well know about the possible topic of the query.

References

1. Blei, D. M., Lafferty, J. D., A correlated topic model of science. *The Annals of Applied Statistics*, 1(1), pp. 17–35 (2007)
2. Blei, D.M., Ng, A.Y., Jordan, M.I., Lafferty, J., Latent dirichlet allocation, *Journal of Machine Learning Research* (2003)
3. Elango, P., Jayaraman, K., Clustering Images Using the Latent Dirichlet Allocation Model (2005)
4. Horster, E., Lienhart, R., Slaney, M., Image retrieval on large-scale image databases. In *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval* (2007)
5. Greif, T., Horster, E., Lienhart, R., Correlated Topic Models for Image Retrieval. Technical Report TR2008-09, Institut fur Informatik, Universitat Augsburg, July (2008)
6. Kullback, S., Leibler, R.A, On Information and Sufficiency. *Annals of Mathematical Statistics*, 22 (1), pp. 79–86 (1951)
7. Lin, J., Divergence measures based on the shannon entropy. *IEEE Trans. Infor. Theory*, 37, pp. 145–151 (1991)
8. Min, J., Leveling, J, Jones, G. J.F., Document expansion for text-based image retrieval at WikipediaMM 2010. In: *CLEF labs 2010, Cross Language Image Retrieval (ImageCLEF)*, pp. 22–23 September 2010, Padua, Italy (2010)
9. Paredes, R., Large-Scale Text to Image Retrieval Using a Bayesian -Neighborhood Model. *SSPR/SPR*, pp. 483–492 (2010)
10. Putthividhya, D., Attias, H. T., Nagarajan, S. S., Topic Regression Multi-Modal Latent Dirichlet Allocation for Image Annotation. In: *CVPR IEEE 2010*, pp. 3408–3415 (2010).
11. Salton, G., McGill, M., *Introduction to Modern Information Retrieval*. McGraw-Hill (1983)
12. Thomee, B., Popescu, A., Overview of the ImageCLEF 2012 Flickr Photo Annotation and Retrieval Task, *CLEF 2012 working notes*, Rome, Italy (2012)
13. Wang, Y. , Mori, G., Max-margin Latent Dirichlet Allocation for Image Classification and Annotation. *British Machine Vision Conference (BMVC)* (2011)

14. Yiming L., Dong X., Ivor W. T., Jiebo L., Textual Query of Personal Photos Facilitated by Large-Scale Web Data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5), pp. 1022-1036 (2011)