RERANKING WITH CONTEXTUAL DISSIMILARITY MEASURES FROM REPRESENTATIONAL BREGMAN *K*-MEANS

Olivier Schwander and Frank Nielsen
École Polytechnique, Palaiseau/Cachan, France
ÉNS Cachan, Cachan, France / Sony Computer Science Laboratories Inc, Tokyo, Japan

Keywords: Image retrieval, Bregman divergences, Alpha divergences, Clustering, Reranking, Context.

Abstract: We present a novel reranking framework for Content Based Image Retrieval (CBIR) systems based on contextual dissimilarity measures. Our work revisit and extend the method of Perronnin *et al.* (Perronnin et al., 2009) which introduces a way to build contexts used in turn to design contextual dissimilarity measures for reranking. Instead of using truncated rank lists from a CBIR engine as contexts, we rather use a clustering algorithm to group similar images from the rank list. We introduce the representational Bregman divergences and further generalize the Bregman *k*-means clustering by considering an embedding representation. These representation functions allows one to interpret α-divergences/projections as Bregman divergences/projections on α-representations. Finally, we validate our approach by presenting some experimental results on ranking performances on the INRIA `Holidays` database.

1 INTRODUCTION

Our work is grounded in the field of Content Based Image Retrieval (CBIR): given a query image, we search similar images in a large dataset of images. Results are displayed in the form of a rank list where images are ordered with respect to their similarity to the query image. Typical CBIR systems manipulate databases of one million images or more (see (Douze et al., 2009; Jégou et al., 2008; Sivic and Zisserman, 2003) for recent works and (Datta et al., 2008) for a comprehensive survey of the field).

Contextual Similarity measures are a way to algorithmically design new similarity measures tailored to the datasets/queries. The word *context* may encompass different meanings in the literature. On the one hand, it can refer to the transformation of a classical divergence $D(p,q)$ into a local divergence $D'(p,q) = \delta(p)\delta(q)D(p,q)$, where the local distance between two points depends on the neighborhood of these two points. This idea was in particular explored in (Jégou et al., 2007), which uses a conformal deformation of the geometry (Wu and Amari, 2002). On the other hand, the notion of context can also refer to a reranking stage with a similarity measure built on the rank list returned by a CBIR system, as developed in (Perronnin et al., 2009). The goal is not only to improve the retrieval accuracy but also to get an ordering that is close to the intent of the user.

Perronnin's system *et al.* (Perronnin et al., 2009) addresses this problem by building contexts and averaging the distances obtained for each context. In this case, contexts are defined as the centroids of truncated rank lists of growing size. We propose to improve this process by building contexts in a more meaningful way: instead of taking the N nearest neighbors of the query, we cluster the rank list and use the centroids of the clusters as contexts in order to naturally take into account the semantic of the rank list. We then use an averaging process to get a unique similarity score to rerank image matching scores.

Instead of using a classical *k*-means clustering algorithm based on the squared Euclidean distance, we rather introduce a modified clustering algorithm based on α-divergences (see Amari (Amari, 2007; Amari and Nagaoka, 2007)). The family of information-theoretic α-divergences are provably more suited to handle histogram distributions at the core of many CBIR systems (e.g., bag of words). We extend the Bregman *k*-means algorithm introduced by Banerjee *et al.* (Banerjee et al., 2005; Nock et al., 2008).

Finally, we evaluate our clustering and reranking framework on the INRIA `holidays` dataset (Jégou et al., 2008) based on the novel contextual similarity measures.

Schwander O. and Nielsen F. (2010).
RERANKING WITH CONTEXTUAL DISSIMILARITY MEASURES FROM REPRESENTATIONAL BREGMAN K-MEANS.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 118-123
DOI: 10.5220/0002842901180123

2 REPRESENTATIONAL BREGMAN DIVERGENCES

2.1 Definitions

Invariance and Information Monotonicity of α-divergences. We recall the definition of α-divergences (Amari and Nagaoka, 2007) that are defined on positive arrays (unnormalized discrete probabilities) for $\alpha \in \mathbb{R}$ as:

$$D_\alpha(p\|q) = \begin{cases} \sum_{i=1}^d \frac{4}{1-\alpha^2}\left(\frac{1-\alpha}{2}p_i + \frac{1+\alpha}{2}q_i - p_i^{\frac{1-\alpha}{2}} q_i^{\frac{1+\alpha}{2}}\right) & \text{if } \alpha \neq \pm 1 \\ \sum_{i=1}^d p_i \log\frac{p_i}{q_i} + q_i - p_i = \mathrm{KL}(p\|q) & \text{if } \alpha = -1 \\ \sum_{i=1}^d p_i \log\frac{q_i}{p_i} + p_i - q_i = \mathrm{KL}(q\|p) & \text{if } \alpha = 1 \end{cases} \quad (1)$$

This is all the more important that in the heart of CBIR systems, we deal with histograms (e.g., bag-of-words) that are considered as multinomial probability distributions. Therefore, we need a distribution measure D to calculate the dissimilarity of multinomials $D(p(x;\theta_p)||p(x;\theta_q))$ where θ_p and θ_q are the histogram distributions. Symmetrized α-divergences $S_\alpha(p,q) = \frac{1}{2}(D_\alpha(p\|q) + D_\alpha(q\|p))$ belong to Csiszár's f-divergences and therefore retain the information monotonicity property.

From the pioneering work of Chentsov (Chentsov, 1982), it is known that the Fisher-Rao riemannian geometry (with the induced Levi-Civita connection) and the α-connections are the *only* differential geometric structures that preserve the measure of probability distributions by reparameterization. We consider the α-divergences that are a proper sub-class of Csiszár f-divergences that satisfy both reparameterization invariance (i.e., $D(p(x;\theta_p)||p(x;\theta_q)) = D(p(x;\lambda_p)||p(x;\lambda_q))$ for $\lambda_x = f(\theta_x)$ where f is a bijective mapping) and information monotonicity (Csiszár, 2008): $D(p(x;\theta_p)||p(x;\theta_q)) \geq D(p(x;\theta'_p)||p(x;\theta'_q))$ for θ' a coarser partition of the histogram. That is, if we merge bins θ into coarser histograms θ', the distance measure should be less than the distance by considering the higher-resolution histograms.

Bregman Divergences. Given a strictly convex and differentiable function $F : \mathbb{R}^d \to \mathbb{R}$, we define the Bregman divergence associated with the generator F as:

$$B_F(p\|q) = F(p) - F(q) - \langle p - q, \nabla F(q)\rangle \quad (2)$$

The generator $F(x) = x^\top x = \sum_{i=1}^d x_i^2$ yields to the squared Euclidean distance. Using the Shannon negative entropy ($F(x) = \sum_{i=1}^d x_i \log x_i$) we get the well-known Kullbach-Leibler (KL) divergence.

2.2 Representation Function

Nielsen and Nock (Nielsen and Nock, 2009) showed that α-divergences (but also β-divergences (Mihoko and Eguchi, 2002)) are representational Bregman divergences in disguise. Let's consider *decomposable* Bregman divergences:

$$B_F(p\|q) = \sum_{i=0}^d B_F(p_i\|q_i) \quad (3)$$

With a slight abuse of notation, we denote its separable generator F as $F(x) = \sum_{i=0}^d F(x_i)$. We call representation function a strictly monotonous function k that introduces a (possibly non-linear) coordinate system $x_i = k(s_i)$ where each s_i comes from the source coordinate system. This mapping is bijective since k is strictly monotonous and $s_i = k^{-1}(x_i)$. We have the following Bregman generator:

$$U(x) = \sum_{i=1}^d U(x_i) = \sum_{i=1}^d U(k(s_i)) = F(s) \quad (4)$$

where $F = U \circ k$.

The class of α-divergences are representational Bregman divergences for

$$U_\alpha(x) = \frac{2}{1+\alpha}\left(\frac{1-\alpha}{2}x\right)^{\frac{2}{1-\alpha}}, \quad k_\alpha(x) = \frac{2}{2-\alpha}x^{\frac{1-\alpha}{2}} \quad (5)$$

Notice it turns out that F may not be strictly convex (Nielsen and Nock, 2009) ($U_\alpha \circ k_\alpha$ is linear) although U always is *strictly* convex.

2.3 Contexts from α k-means: α-centroids

Like (most of) the Bregman divergences, α-divergences are not symmetrical. This yields two different ways of defining centroids: the left-sided centroid c^L and the right-sided centroid c^R:

$$c^R = \arg\min_{c\in\mathcal{X}} \sum_{i=1}^n B_{U,k}(p_i\|c) \quad (6)$$

$$c^L = \arg\min_{c\in\mathcal{X}} \sum_{i=1}^n B_{U,k}(c\|p_i) \quad (7)$$

Closed-form formulas are given in (Nielsen and Nock, 2009):

$$c^R = n^{-\frac{2}{1-\alpha}} \left(\sum p_i^{\frac{1-\alpha}{2}} \right)^{\frac{2}{1-\alpha}} \quad (8)$$

$$c^L = n^{-\frac{2}{1+\alpha}} \left(\sum p_i^{\frac{1+\alpha}{2}} \right)^{\frac{2}{1+\alpha}} \quad (9)$$

2.4 Clustering with Representation Functions

Banerjee *et al.* (Banerjee et al., 2005) showed that the classical clustering algorithm k-means generalizes to and only to Bregman divergences. Using the representational framework of section 2.2, we extend their algorithm to the α-divergences by plugging the representation function. This leads to the algorithm 1, which is nearly identical to the classical one except for the centroid computation part. (Symmetrized α-divergences S_α are handled implicitly by two potential functions, similarly to (Nock et al., 2008).)

Algorithm 1 Representational Bregman k-means.

Require: A set $\mathcal{X}$ of n points x_i of $\mathbb{R}^d$, a number of clusters k, a Bregman representational divergence $B_{U,k}$

Ensure: A hard partitioning $\{\mu_i\}_{1\leq i\leq k}$ of $\mathcal{X}$ which is a local minimizer of the loss function $\sum_{h=1}^{k} \sum_{x_i \in \mathcal{X}_i} B_{U,k}(x_i, \mu_h)$

Choose k points μ_i (with k-means++ initialization method (Nock et al., 2008))

repeat

{Assignment step}

Set $\mathcal{X}_h \leftarrow \emptyset$ for $1 \leq h \leq k$

for $i = 1$ to n **do**

$h \leftarrow \arg\min_{h'} B_{U,k}(x_i || \mu_{h'})$

Add x_i to $\mathcal{X}_h$

end for

{Relocation step}

for $h = 1$ to k **do**

$\mu_h \leftarrow k^{-1} \left(\sum_{i=1}^{n} k(x_i) \right)$

end for

until convergence

Return $\{\mu_1, \ldots, \mu_k\}$

3 CONTEXTUAL DISSIMILARITY MEASURES

3.1 Definition

Perronnin *et al.* (Perronnin et al., 2009) introduces a framework to improve the retrieval performance of the bag-of-word CBIR systems. The following is defined for any arbitrary divergence f. Let's first define a function Φ:

$$\Phi_f(\omega; q, p, u) = f(q, \omega p + (1-\omega)u) \quad (10)$$

The contextual dissimilarity will be defined as the following minimization problem[1]:

$$\mathrm{cs}_f(q, p|u) = \arg \min_{0\leq\omega\leq 1} \Phi_g(\omega; q, p, u) \quad (11)$$

This minimization is equivalent to searching for the generalized Bregman projection of the point q on the (euclidean) line segment $[p,u]$ (see Figure 1, $p_\perp$ denotes the usual Euclidean projection and $p_\star$ the Bregman projection) and thus can be solved using a simple convex minimization algorithm.

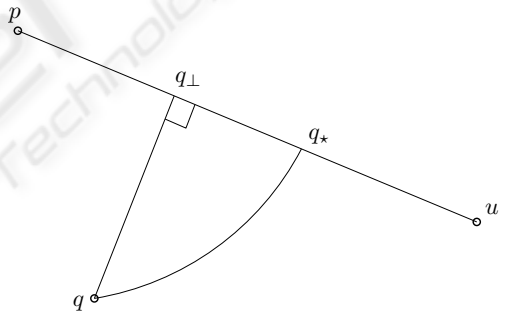


Figure 1: Contextual similarity and Bregman projection.

One can also use the symmetrized version of this measure. In this case, we take:

$$\Phi_f(\omega; q, p, u) = f(q, \omega p + (1-\omega)u) + f(q, \omega p + (1-\omega)u) \quad (12)$$

4 RETRIEVAL PROCESS

4.1 Retrieval

We choose to use a retrieval system based on GIST descriptors as described in Jégou *et al.* (Jégou et al.,

[1]Here, we use dissimilarity instead of similarity, so we get a min instead of the max of Perronnin *et al.*

2007). GIST descriptors are global image descriptors (introduced in (Oliva and Torralba, 2006)) that allow quite good performances scores (even if inferior to state-of-the-art bag of words approaches, see (Jégou et al., 2008)) with *reduced memory* footprint and *high speed* that allow systems to scale well on large datasets.

The GIST CBIR framework works as follows:

1. Given a query image, compute its GIST descriptor.
2. Search for N-nearest neighbors of the query in the dataset (where N is the size of the short list).
3. Display the neighbors ordered with respect to their distances to the query.

4.2 Reranking

The contextual dissimilarity takes place as a reranking step after the use of a classical retrieval system. It was shown empirically in (Perronnin et al., 2009) that images that are not in the short list have very low chance to be in the new N-neighborhood.

Given a short list of size N, the algorithm is the following:

1. Take only the first k elements (i.e. the k nearest neighbors of the query).
2. Estimate a context using the selected points by computing their centroids.
3. For all elements of the short list, compute $\mathrm{cs}_f(q, p_i|u_k)$.
4. Rerank the list according to these scores.

4.3 Reranking with Multiple Contexts

The previous algorithm uses only one context which may be not sufficient to capture the information related to contexts. The original framework of Perronnin *et al.* (Perronnin et al., 2009) used truncated rank lists to build the contexts. The full dissimilarity is computed by doing a weighted average of the dissimilarities with context built using the centroid of growing size short lists. This approach leads to good experimental results. However, the truncated rank list is not the better method to capture the meaning of clusters. As depicted in Figure 2, a truncated rank list may group images from different groups of similar images.

So instead of using the k nearest neighbors of the query to define a context, we cluster the rank list in order to get meaningful contexts which best describe the query. We then propose the following average scheme:

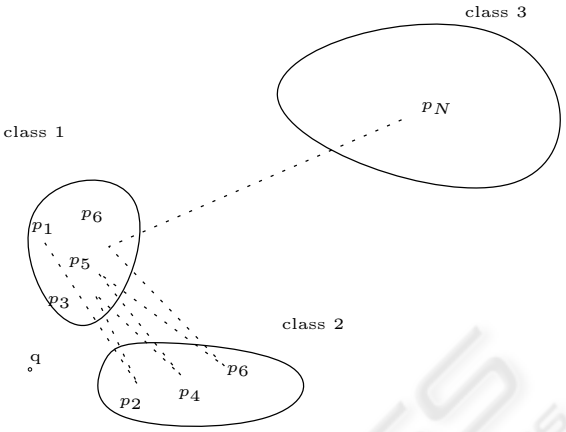


Figure 2: Different contexts and potential rank lists.

$$\mathrm{cs}_f(q,p) = \sum_{u \in \mathcal{U}} \mathrm{cs}_f(q, p\,|\,u) \tag{13}$$

5 EXPERIMENTS

5.1 Experimental Setup

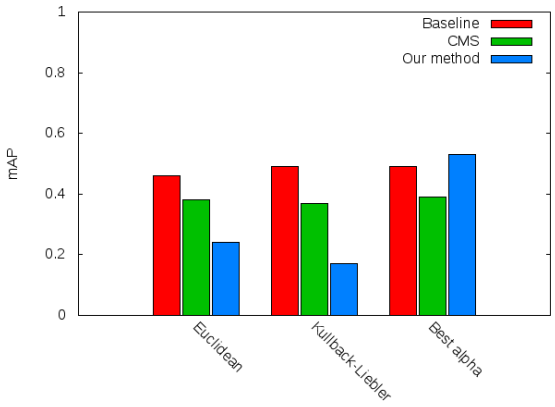
Dataset. We use the Holidays dataset from INRIA: this set was introduced to evaluate state-of-the-art framework (Jégou et al., 2008). It contains 1491 personal holiday photos with approximately 500 image groups. Thanks to INRIA, this dataset is publicly available on the author's website[2].

Baseline. We first report the results for a simple ranking system based on GIST global descriptors (see (Douze et al., 2009)). We use α-divergences as dissimilarity measures.

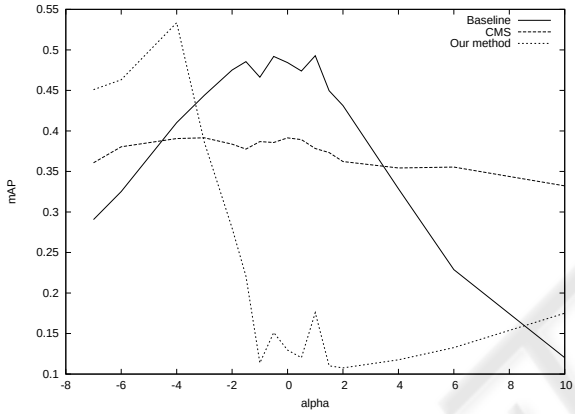
Contextual Measure of Dissimilarity. We next present results for the Contextual Measure of Dissimilarity of (Perronnin et al., 2009). The results are not directly comparable with the original paper since we use global descriptors instead of a bag-of-features approach. Moreover, we do not only use Kullbach-Leibler divergence to do the ranking but also α-divergences.

Evaluation. The retrieval accuracy is measured in terms of mean average precision (mAP) which is a very common measure in the information retrieval field.

[2] http://lear.inrialpes.fr/people/jegou/data.php



(a) Comparison between different reranking methods



(b) Impact of the α parameter

Figure 3: Experimental results. *Baseline* is the raw rank list. *CDM* is the Contextual Dissimilarity Measure of (Perronnin et al., 2009)

5.2 Results

We compare on the Figure 3a the results we get for: a rank list without any reranking (baseline), the (Perronnin et al., 2009) contextual dissimilarity measure (CDM) and our method. Results are shown for the Euclidean divergence, the Kullbach-Leibler divergence and the α-divergence with the α which gave the best mAP score.

We see that our reranking method behaves badly for the Euclidean and Kullbach-Leibler divergence but we manage to outperform these two divergences with a well-chosen α.

In the Figure 3b, we study the impact of the α parameter for the three different methods (baseline, Perronnin and ours). The influence is not so big on (Perronnin et al., 2009) CDM but the scores really depend on the α for the two other methods. Moreover, the behavior is completely different for our method and for the Perronnin one's: our method reaches an optimum near $\alpha = -4$ and is not symmetrical whereas the CDM curve is symmetrical and centered on $\alpha = 0$.

6 CONCLUSIONS

We present a new family of clustering algorithms based on *k*-means and α-divergences. We recall how representational functions can be used to map α-divergences to the space of the well-known Bregman divergences. Using this mapping, we show that we can adapt the Bregman *k*-means algorithm in order to build an α *k*-means clustering algorithm.

We then focus on contextual similarity measure by using this family of *k*-means algorithms to build contexts by clustering the rank list given by a traditional (that is to say, not contextual) retrieval system.

Using α-divergences with a well-chosen α parameter and our cluster based contextual dissimilarity measure, we are able to outperform other contextual similarity measures. Since the choice of α is critical for the quality of the results, we can conclude that the α divergences are a very interesting family of dissimilarity measures.

ACKNOWLEDGEMENTS

We thank Hervé Jégou (INRIA) and Shun-ichi Amari (Brain Science Institute) for insightful email exchanges considering respectively the reranking methodologies and the α-divergences. We also thank the referees for their helpful observations.

REFERENCES

Amari, S. (2007). Integration of stochastic models by minimizing α-divergence. *Neural computation*, 19(10):2780–2796.

Amari, S. and Nagaoka, H. (2007). *Methods of information geometry*. AMS.

Banerjee, A., Merugu, S., Dhillon, I., and Ghosh, J. (2005). Clustering with Bregman divergences. *The Journal of Machine Learning Research*, 6:1705–1749.

Chentsov, N. (1982). *Statistical Decision Rules and Optimal Inferences*. Trans. of Math. Monog., n 53.

Csiszár, I. (2008). Axiomatic characterizations of information measures. *Entropy*, 10(3):261–273.

Datta, R., Joshi, D., Li, J., and Wang, J. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*

Douze, M., Jégou, H., Singh, H., Amsaleg, L., and Schmid, C. (2009). Evaluation of GIST descriptors for web-scale image search. In *International Conference on Image and Video Retrieval*. ACM.

Jégou, H., Douze, M., and Schmid, C. (2008). Hamming embedding and weak geometric consistency for large scale image search. In *European conference on computer vision*, pages 304–317. Springer.

Jégou, H., Harzallah, H., and Schmid, C. (2007). A contextual dissimilarity measure for accurate and efficient image search. In *Conference on Computer Vision & Pattern Recognition*.

Mihoko, M. and Eguchi, S. (2002). Robust blind source separation by beta divergence. *Neural computation*, 14(8):1859–1886.

Nielsen, F. and Nock, R. (2009). The dual Voronoi diagrams with respect to representational Bregman divergences. In *International Symposium on Voronoi Diagrams (ISVD)*.

Nock, R., Luosto, P., and Kivinen, J. (2008). Mixed bregman clustering with approximation guarantees. In Daelemans, W., Goethals, B., and Morik, K., editors, *ECML PKDD (2)*, volume 5212 of *Lecture Notes in Computer Science*, pages 154–169. Springer.

Oliva, A. and Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, 155:23.

Perronnin, F., Liu, Y., and Renders, J. (2009). A family of contextual measures of similarity between distributions application to image retrieval. In *CVPR09*, pages 2358–2365.

Sivic, J. and Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. In *Proc. ICCV*, volume 2, pages 1470–1477. Citeseer.

Wu, S. and Amari, S. (2002). Conformal transformation of kernel functions: A data-dependent way to improve support vector machine classifiers. *Neural Processing Letters*, 15(1):59–67.