

# THE DIGITAL IMAGE CLASSIFICATION BASED ON CALCULATION OF RENYI DIVERGENCE

V. D. Sergeev

Saint-Petersburg State University, Math.& Mech. faculty,

Email: vlad.sergeev.spbgu@gmail.com

## Abstract

*The method of analysis and classification of images based on the calculation of alpha-divergences (or Renyi's divergences) is considered. For each image we construct a discrete measure distribution and finite sequence of its direct multifractal transforms. The resulting vector is a characteristic of the image structure. For two images, the  $\alpha$ -divergences between these vectors are calculated. We perform experiments and provide results with several classifiers that use these vectors as training and testing datasets.*

## 1. INTRODUCTION

Image analysis and classification play a special role in biology and medicine because this is the one of the main ways to observe the system under investigation. The results of observations can be saved as digital images.

The characteristics of a digital image can naturally be determined in terms of intensities of its constituent pixels. One of the well-known methods consists of dividing the image into cells and calculating certain measures defined on this partition. The easiest way to determine the measure of a cell is to calculate the sum of its pixel intensities, but it is also possible to apply different filters. The measure for color images may be defined for any of the palette component.

The measure distribution in cells allows the use of entropy characteristics for the analysis and classification of images. The natural characteristic is the entropy of the partition [3], which is defined as  $-\sum_i \mu_i \ln \mu_i$ , where  $i$  — is the number of elements

in the partition, and  $\mu_i$  is the measure of  $i$ th cell. The entropy of a partition is sometimes called the entropy of a probability vector or the Shannon entropy. This entropy is a special case of the class of Renyi entropies — a set of characteristics depending on a real parameter. The Renyi entropies are widely used both in image analysis, and as objective functions in solving problems of isolated extremes [2,4].

The calculation of entropy allows us to obtain certain digital characteristic of an image. At the same time, it should be noted that the entropy value does

not depend on the order of the component of the probability distribution vector. Therefore images with different structures may have different distribution vectors, but if one vector is a permutation of components of another, the entropies of these vectors will be the same.

To find the difference in the structure of the compared images one may calculate Renyi's divergence. The Kulbak-Leibler divergence is the most commonly used. It is not enough to obtain one numerical characteristic for image classification, so the methods which allow us to obtain several numerical characteristics are more preferable. Resulting vectors allow us to find rather subtle differences in the structure of images.

In this paper, we consider a method of obtaining vector characteristics — vectors containing Renyi divergences between initial measures and their direct multifractal transforms, and the application of the method to image classification. We used three well-known classification models: support vector machine, decision tree and random forest. The classification was performed for some classes of medical images.

The work has the following structure. The next section describes Renyi entropy and direct multifractal transform. In the section 3 we give brief description of the used classification methods. The next section contains the results of experiments performed for the class of images of kidney tissue.

## 2. MAIN DEFINITIONS

This image comparison method is based on calculation of the Renyi divergence (or  $\alpha$ -divergence).

The method consists in obtaining a certain discrete measure for each image, which is calculated for a given partition. The measure of each cell is defined as the sum of pixel intensities. Then the measure is normed that results in obtaining the probability measure distribution. We also consider a finite sequence of direct multifractal transforms of the initial distribution (the number of such transforms is determined empirically). The  $\alpha$ -divergence vector is calculated between the resulting sets of measures for two compared images, which we consider as a characteristic of the similarity-difference of image structures

For two given distributions  $\mathbf{p} = \{p_i\}$  and  $\mathbf{q} = \{q_i\}$  which characterize two compared images, the Renyi divergences for the variable parameter  $\alpha > 0$ ,  $\alpha \neq 1$  are defined as

$$D_\alpha(p, q) = \frac{1}{\alpha - 1} \ln \sum_1^n p_i^\alpha q_i^{1-\alpha} \quad (1)$$

This value is non-negative, and the divergence is a non-decreasing function of  $\alpha$ .

In the particular case when  $\alpha = 1$ , this transformation is determined by the formula:

$$D_1(p, q) = \sum_1^n p_i \ln \frac{p_i}{q_i} \quad (2)$$

and is called the Kullbak-Leibler divergence.

For a given distribution  $\mathbf{p} = \{p_i\}$ , the direct multifractal transform of the initial distribution is determined by the formula:

$$f_k(p) = \frac{p^k}{\sum_i p_i^k} \quad (3)$$

where  $k$  — is a real value.

Direct multifractal transform acts on a set of discrete probability measures corresponding to the images. These transforms form a group, and the set of measures decomposes into disjoint classes of transitivity. Each class contains some initial measure corresponding to the given image, and its subsequent transformations. In fact, the resulting class of measures characterizes the selected image in a certain way.

Experiments show that this method allows us to divide the images into groups according to their actual (expert) classification [1].

### 3. CLASSIFICATION MODELS

The task of image classification is of the great importance in many subject areas. Usually to solve this problem supervised, unsupervised and deep learning models are considered. We select supervised models because they are easier to measure and verify. In this work 3 classifiers are used: support vector machine (SVM), decision tree and random forest.

SVM is a binary classification algorithm. Among the most large-scale problems that were solved using SVM (and its modified implementations) are the display of advertising banners on sites, the recognition of sex on the basis of photography and the splicing of human DNA.

In classification tasks decision tree model assumes that the predicted result is the class to which the data belongs. The model is simple in understanding and interpretation and allows evaluating the model using statistical tests. This gives us an opportunity to estimate the reliability of the model.

Random forest model is a composition of decision trees. The final result of a random forest classification will be the class for which the majority of the trees voted. We suppose that one tree has one voice. If a model with 500 trees is created for binary classification problem, among which 100 trees point to the zero class, and the remaining 400 point to the first class, the model will predict the first class as a result. This model has high parallelism and scalability.

All implemented models are developed using Java 8 with Apache Spark MLlib algorithms[5].

SVM is configured to use stochastic gradient descent with 500 iterations and different seed values, that optimize model training. L1 and L2 regularization method is used to prevent overfitting of model.

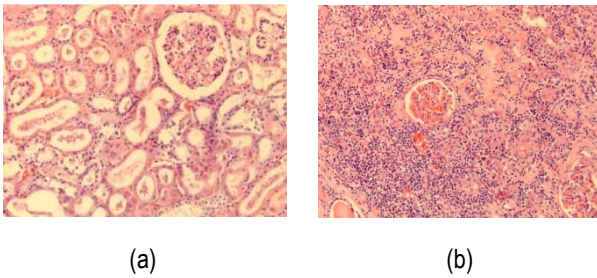
Decision tree has a lot of parameters to configure. In our experiments we use two impurity measures for classification: Gini impurity and entropy. Maximum depth of decision tree is 30.

Random forest also has a lot of parameters and most of them are equal to decision tree model, so we use the same values for them. We configure 200 trees and different seeds for tree generation algorithm.

## 4. RESULTS OF EXERIMENTS

We performed several experiments using two classes of biomedical images. Images were split by experts to groups according to the information they describe: healthy (14 images) and diseased kidney (15 images). Calculation of divergence is not symmetric operation, so for two images we calculated divergence vectors twice. We calculated Renyi divergence vectors for each pair from the group. For each class of images around 200 feature vectors were obtained. Then one group was marked as 1, and the other as 0.

Renyi divergence was calculated for  $\alpha=\{0.25, 0.5, 1, 2\}$  in grayscale palette. Each image was partitioned into cells with size 100x100 pixels for probability distribution calculation. For each cell we calculated measure as the sum of pixels intensity divided by the sum of intensities for whole image. Multifractal transforms were calculated with power from 1 to 10, which means that each divergency vector has the size 10.



**Fig. 1.** Examples of kidney images where (a) is healthy image and (b) with disease

We got 4 groups of vectors and then performed models calculation for each group of vectors according to  $\alpha$  values. Then we combined all vectors for healthy kidneys and did the same for kidneys with diseases, and train models over these two big datasets.

Before running dataset (union of positive and negative vectors) over each classifier, dataset was divided in two parts: 70% of dataset was used for training model and 30% for testing of the trained model.

### 4.1. SVM

Main estimation for SVM classification model is area under ROC-curve. Seed value is used for SGD optimization, so that we could obtain same results on the same dataset.

Seed value / $\alpha$	0.25	0.5	1	2	All
64	0.5	0.5	0.9	0.84	0.68
97	0.5	0.5	0.88	0.84	0.71
34	0.5	0.5	0.9	0.85	0.71

**Table 1.** SVM classification results for L1 regularization

Seed value / $\alpha$	0.25	0.5	1	2	All
52	0.74	0.67	0.81	0.76	0.67
68	0.86	0.85	0.91	0.88	0.71
38	0.84	0.71	0.87	0.84	0.69

**Table 2.** SVM classification results for L2 regularization

As we can see from results above L2 regularization provides better classification results. Sparsity refers to that only very few entries in a matrix (or vector) is non-zero. L1-norm has produces many coefficients with zero values or very small values with few large coefficients, that can provide worse result.

### 4.2. Decision tree

Quality of the decision tree model can be measured by the frequency of an error which occurs in classification of model. It is calculated as a number of negative predicted test vectors divided by a total number of test vectors.

$\alpha=0.25$	$\alpha=0.5$	$\alpha=1$	$\alpha=2$	All
0.2	0.09	0.26	0.27	0.23
0.19	0.10	0.24	0.26	0.23
0.18	0.15	0.29	0.31	0.24

**Table 3.** Decision tree classification results for Gini impurity

$\alpha=0.25$	$\alpha=0.5$	$\alpha=1$	$\alpha=2$	All
0.23	0.12	0.18	0.2	0.26
0.19	0.13	0.17	0.21	0.24
0.19	0.07	0.19	0.24	0.24

**Table 4.** Decision tree classification results for entropy impurity

Classification results can be improved with better  $\alpha$  values. In general the both impurity measures provide the same averaged result.

### 4.3. Random forest

Quality of the random forest model can be estimated by the same way as for decision tree. To improve classification accuracy over a single decision tree, the individual trees in a random forest need to differ. This difference is achieved by introducing randomness in the generation of the trees. The randomness is influenced by the seed, and

what is the most important is that using the same seed should always generate the same result.

Seed value / $\alpha$	0.25	0.5	1	2	All
54	0.18	0.1	0.22	0.27	0.24
57	0.24	0.08	0.22	0.18	0.23
90	0.17	0.15	0.2	0.25	0.22

**Table 5.** Random forest classification results for Gini impurity

Seed value / $\alpha$	0.25	0.5	1	2	All
94	0.19	0.11	0.18	0.22	0.22
61	0.2	0.17	0.21	0.21	0.23
70	0.18	0.09	0.16	0.22	0.22

**Table 6.** Random forest classification results for entropy impurity

As we can see, decision tree model classification results could be improved with better selection of  $\alpha$  values. But in average both measures provide the same averaged result.

## 5. CONCLUSION

There are a lot of parameters that we can tune in Renyi divergence calculation. We can increase the power of multifractal transform, select  $\alpha$  from widely range for specific images and change cell size for

probability distribution calculation. Parameters tuning depends on the class of images and is performed empirically.

Experiments showed that SVM model has better results in classification divergence vectors for  $\alpha=\{1, 2\}$ . For  $\alpha=\{0.25, 0.5\}$  decision tree and random forest have better classification, but for union of all feature vectors obtained for  $\alpha=\{0.25, 0.5, 1, 2\}$  random forest has better estimation of classification quality.

Thus, using Renyi divergence provides good results in specified configuration for medical images, and in conjunction with other features may increase classification models output quality.

## References

- [1] Ampilova N.B., Sergeev V.D., Soloviev I.P. Digital image analysis based on direct multifractal transform. Humanities and Science University Journal 19(2016), c.23-32, <http://submit.uni-journal.ru/article/8850>
- [2] N. Ampilova, I. Soloviev. On Application of Entropy Characteristics to Texture Analysis, WSEAS Transactions on Biology and Biomedicine, ISSN / E-ISSN: 1109-9518 / 2224-2902, Volume 11, 2014, Art. #25, pp. 194-202.
- [3] B. Hasselblat, A. Katok. Introduction to the Modern Theory of the Dynamical Systems, M., Factorial, 1999
- [4] V. Vstovsky. Elements of information physics, M., 2002.
- [5] <https://spark.apache.org/docs/latest/ml-guide.html>