Data Field-based Support Vector Machine for Image Classification

Yi Lin

Computer School, Wuhan University, Wuhan, Hubei, 430072, P.R. China

linyi777@live.cn

Keywords: Image Classification, Data field, topological Potential, Diffusion distance, SVM

Abstract. In computer vision, object recognition is still a challenge. In this paper, a new method based on data field is proposed for image object classification with color histograms and diffusion distances. Among them, the topological potential is used to select the optimal parameters. The experimental results show that the proposed method has better accuracy and shorter run time than four common kernels. It not only overcomes the drawbacks of the existing parameter selection method, but also coincident with Vapnik's theory, which theoretically guarantees the generalization of learning machines.

1. Introduction

Recognition is a class of classical problem in computer vision, on which many specialized tasks are based. In ICCV 2013, the papers about recognition accounted for nearly 30%. Being a valuable technology, object recognition has been deeply infiltrated the production life of all walks of life, like Remote sensing, robots, autonomous cares, android eyes, image panoramas, etc.

Image Classification has been a challenge in computer vision over multiple decades. Generally, the various recognition problems related to a definite application can be solved and implemented in different ways. Total approaches can be divided into the following categories: based on CAD-like object models, appearance-based methods, feature-based methods [1, 2], genetic algorithms[3] and other approaches. In this paper, we use color histogram[4] as an image representation; then, because of support vector machine(SVM)[5]'s high generalization performance, we choose it as the classification model, which can successfully avoids the curse of dimensionality. Among them, we use the data field function and topology potential for custom kernel and parameter selecting. A series of experiments shows that the proposed approach does much better than common kernels.

This paper follows the automatically identification process of the best kernel and its parameter settings. Section 2 provides a brief background introduction of data field. Section 3 describes how to construct data field kernel and selecting its parameters. Section 4 explores this method's advantages and limitations, and proposed improvements in the future. Section 5 gives the paper's summary.

2. Method

2.1 Color Histogram and Diffuse Distance

In computer vision, the color histogram[6][6][10] represents the colors' distribution of an image. It is a kind of simple and fast low-level representation for digital image. The color histogram is produced by discretization of the image's colors into several bins and counting the number of image pixels in each bin. The spatial features may be lost in computation process, but it guarantees the full conversion and rotational invariance of image.

Diffusion distance[7] is a novel cross-class distance based on histogram descriptors, which understand the difference between histograms by a heat diffusion process in physics. Diffusion distance is not only robust to any distortions such as image deformation, light changes and noise that usually causes problems for HBLDS, but also have a linear time complexity.

If a m-dimensional histogram is denoted by $h_1(x)$, then $h_1(x)$, $h_2(x)$ represents two different histograms. The diffusion distance is defined as $K(h_1, h_2) = \sum_{i=0}^{L} k(|d_1(x)|)$, where $d_0 = h_1(x) - h_2(x)$, $d_1 = [d_{l-1}(x) \times \phi(x, \sigma)] \downarrow_2$, $(l = 1, \dots, \sigma)$. The notation $'\downarrow_2'$ denotes half size down-sampling. L is the number of pyramid layers and σ is the constant standard deviation of Gaussian pyramid. Owing to the low computation load and good performance, paradigm L_1 is used to compute $k(\cdot)$.

2.2 Data Field

Inspired by modern field theory, data field[8] introduces the interaction between material particles and the description way of field into abstract data space. Let $D = \{x_1, x_2, \dots, x_n\}$ be a data set with n objects in space χ, where $x_i x = (x_{i1}, x_{i2}, \dots, x_{ip})', (i = 1,2, \dots, n)$, then each object is equivalent to a particle with some quality in a p-dimensional space. In whole space, there is a data field around and any object with field receives the associative action of other objects. In principle, any function morphology which meets above conditions can be used to define data field. Reference to the gravitational field and nuclear field, it can give the function morphology of quasi gravity field (1) and quasi nuclear field (2), as below:

$$\varphi_x(y) = \frac{m}{1+\left(\frac{\|x-y\|}{\sigma}\right)^k}. \tag{1}$$

$$\varphi_x(y) = m \times e^{-\left(\frac{\|x-y\|}{\sigma}\right)^k}. \tag{2}$$

Where, $m \geq 0$ represents the strength of field source, which can be review as the quality of data object; $\sigma \in (0, \infty)$, named impact factor, is used to control the interaction range between objects; and $k \in N$ is the distance index.

In physics, the potential is defined as the work done by field force. It means that the work done by moving a unit particle from one place to reference within field. So, the potential of data field is defined as follows, which is usually a single value function of spatial position. If the data set $D = \{x_1, x_2, \dots, x_n\}$ contains n objects in the space $\Omega \subseteq R^p$, then the potential value of the data field at any place $x \in \Omega$ in space can be expressed as:

$$\varphi(x) = \varphi_{\mathcal{D}}(x) = \begin{cases} \sum_{i=1}^{n} m_i \times \varphi_{gi}(x - x_i) \\ \sum_{i=1}^{n} m_i \times \varphi_{ni}(x - x_i) \end{cases}. \tag{3}$$

where $\varphi_{gi}(x - x_i)$ is the quasi gravitational potential value of x, $\varphi_{ni}(x - x_i)$ corresponds to a monodromic potential value generated by quasi nuclear field; $m_i \geq 0$ is the mass of object x_i, which is normalized, i.e., $\sum_{i=1}^{n} m_i = 1$. Last but not the least, the value of topology potential has nothing to do with the existence of particle at this position.

2.3Kernel Method

The theory of kernel function[9] has a very long history. Mercer's theorem[10] can be traced back to 1909.The research of reproducing kernel Hilbert space (RKHS)[11] was emerged in 1940s. As early as 1964, Aizermann and other[12] introduced it into the field of machine learning in the study on potential function method. Until 1990s, it was used in support vector machine by Boser, Guyon and Vapnik[5] successfully, whose importance began to be taken seriously.

In the beginning of SVM research, people more concerned about the design algorithm based on kernel functions. Later, it was recognized that one of keys to improve the performance of SVM is to design an appropriate kernel for a given problem. In theory, a kernel is often transformed from classic kernels based on the basic property of kernel function. In the principle of structural risk minimization, the SVM classifier design process may take two steps. Frist of all, it can choose an appropriate subset of functions that it has the best classification ability for problem; then, it chose a discriminant function from this subset to make the experiment risk minimization.

3. Experiments and Results

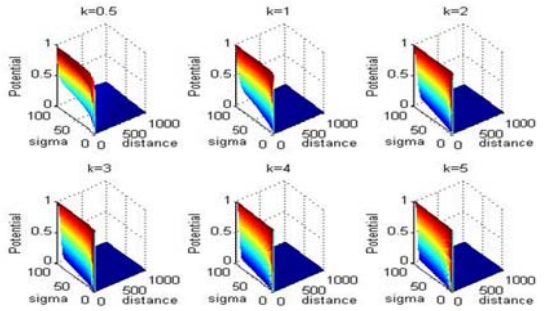
3.1 Function Morphology Selecting

The SVM's design framework is relatively simple, whose performance mainly depends on kernels and their parameters. The kernel function implicitly determines the form of mapping function, which is more important to determine the corresponding feature space of original data [13].

In order to find the more appropriate kernel, we did a series of experiments to evaluate the general performances of different kernels. For SVM with ten different kernels, 1000 random tests were tested respectively, which included four common kernel functions (linear kernel, polynomial kernel, RBF

kernel and sigmoid kernel), the pseudo-gravitational potential functions with three different K (K=0.5，K=1 and K=2) and the pseudo-nuclear potential functions with different K (K=0.5，K=1 and K=2).

In these experiments, the SVM error penalty parameter C was set to 100, which is "large" enough in most cases. The purpose is to enforce the full separability for all kernels except linear kernel. For the reset kernels, the gamma function of kernel function is set to the reciprocal of the number of attributes, which is the recommended defaults of Libsvm official. According to Schoenberg's conclusions [14], if $0 \leq p \leq 2$, function $f(x - x_i) = exp\{-a|x - x_i|^p\}(a > 0)$ is positive definite and obeys the "Mercer Condition". So the K should be a value in the range $0 \leq k \leq 2$.Figure 1, Table 1 and Table 2 respectively shows the box plot of classification accuracy and time cost for modeling.



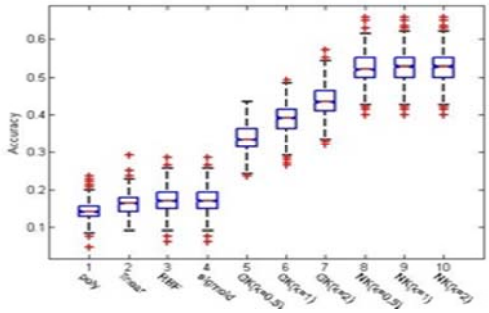


Fig. 1 Box Plot of Classification Accuracy with Different Kernel. Fig. 2 The Surface of Quasi Nuclear Field with Different K

Table 1. Time Cost for Modeling with four common Kernel

Kernel	Poly	Linear	RBF	Sigmoid
Modeling Time (S)	1.7492e+05	1.1498e+05	9.1467	3.8804

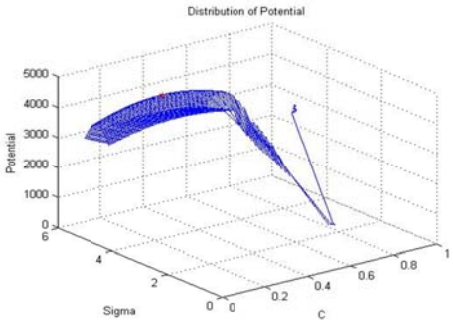
Table 2. Time cost for Modeling with Data Field Kernel

kernel	Quasi Gravity Field			Quasi Nuclear Field		
	K=0.5	K=1	K=2	K=0.5	K=1	K=2
Time(S)	2.5709	2.6084	2.6613	4.9929	2.5859	2.5450

It is clear that there are 3 echelons in classification accuracy. The first echelon includes four common kernels, whose performance is the most inferior in three echelons. The middle echelon is the SVM with quasi gravity kernels, whose accuracy increases with the value of K. The top echelon is the SVM with quasi nuclear kernels, whose classification accuracy do not have too great relationship with K. In addition, the modeling time based on linear kernel or polynomial kernel is much longer than others, which is not suitable for practical applications. Other kernels' cost basically is in the same order of magnitude. In summary, the quasi nuclear kernel is the best choice from the angle of classification accuracy and computation cost.

3.2. Index Identification

In figure 2, there are only small differences between quasi nuclear fields with different k. This is because of the nature of quasi nuclear kernel and morphological of discriminant function $f(x) = sgn(\sum_{i=1}^{N} \alpha_i^0 \, y_i \mathrm{K}(x_i, x) + b)$. Figure 3 shows function surface of quasi nuclear kernels with different k. It is easy to find that the index of quasi nuclear field has little impact on its function value, especially $k \geq 2$. Despite $K(x_i, x)$ varies with different indices, the sign function $f(x)$ may be the same. As shown in figure 4, the different k makes different superplane, but the classification result is indeed the same. In view of simple calculation and good classification performance, the k is taken as one.



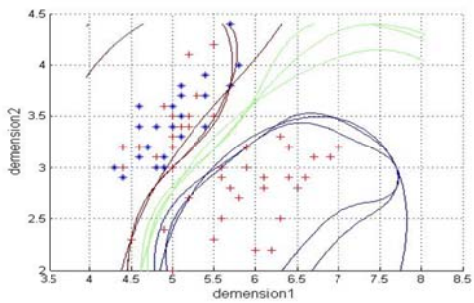


Fig. 3 The Decision Boundaries with Different K. Fig. 4 Topological potential of candidate settings

3.3. Parameter Selecting

The mainstream methods for SVM's parameter selecting include two categories: grid search and heuristic algorithms. The former can get the global optimal solution, but it costs too much in a wide range; the latter mainly refers to the genetic algorithms[15] and particle swarm optimization algorithms[16], etc. They can find the optimal solution without traversing the whole parameter space, but they are easy to fall into local minima.

In this study, we first use the particle swarm algorithm to shrink the parameters' ranges; then gird search is used to refine selecting; finally, we calculated these points' topology potential. The optimal parameters' value is the setting with maximum potential value. To avoid falling into local extreme points, the position variation operation is introduced after updating the velocity and position, which makes partials move with variation in solution space and update particle extreme and global extreme gradually until the iteration conditions is met. The best position which all particles passed is taken as the initial search range of fine gird search.

Unfortunately, there are several settings can reach the top CV accuracy. In most cases, it uses random or the first setting for instead. In my view, each setting is viewed as a particle with certain performance, its surrounding exists a performance field. Any object in the field is subject to the combined action of other objects; the particle with larger potential value has a higher performance, whereas the potential value of particle with low performance is small. So, the topology potential can be used to choose the optimal setting, with $o(n^2)$ cost. Reference to definition of topology potential, the potential value nearby particles with high performance or particle-intensive areas is larger, so the performances of particles nearby the largest topology potential are higher, i.e. the topology potential can be used to reflect the appearance possibility of the best parameter setting. Vividly, the top topology potential must be occurred neared the center position of these settings, which away from individual outliers and the OSH is not too simple or complex. In this way, there will be no "owe-fitting" or "over-fitting", which ensure the generalization performance as far as possible.

3.4. Results on Simulation Dataset

The Caltech 101[17, 18] is a data set of digital images, which is intended to facilitate the researchers of computer vision, especially in recognition, classification, and categorization. It contains a total of 9146 images, split between 101 distinct objects and an additional background category. To simulate real-word classification problem, the 29 colored original categories are merged into seven. Table 3 illustrates the relationship between them.

Taking the instance with random state 2.040387102000000e+09 for example, the specific parameter selection processes unfold as follows. After PSO roughing, the initial parameter range was refined to $c \in (0,1), \alpha \in [1,6]$. In grid fine selection, the step length of parameters is taken as 0.01. A total of 6130 settings reach the highest CV. Figure 4 shows the topological potential of these settings, where the red star marks the setting C =0.31, S=4.58 with top potential value.

Table 3. New categories generated Caltech 101

New	Size	Original category in Caltech101
Fish	57	sea_horse
Reptile	201	crocodile, crocodile_head, hawksbill
Bird	992	Airplanes, flamingo, flamingo_head, ibis
Mammal	1431	Faces, Faces_easy, Leopards, cougar_body, dalmation, gerenuk, hedgehog, kangaroo, okapi, platypus
Invertebrate	84	scorpion
vegetation	285	Bonsai, strawberry, sunflower, water_lilly
abiotic	533	grand_piano, headphone, lamp, pagoda, stapler, watch

4. DISSCUSSION

In this study, I have shown that using data field function and topology potential to construct kernels and select its parameter. The experiments demonstrate that the proposed kernels can do much better than common kernels for a more generic image sets. Besides, the topology potential based on method give a rational physics explanation for parameter selecting. This method not only has a low

time complexity, but also without other prior knowledge and extra parameters. The last but not the least is that the selected parameter is in line with the structure risk theory. This method can be used in many processes of automation production like image retrieval, detection, segmentation, action tasks etc. The data field kernel could be tried in many other applications, if the specific industry background can be integrated, it may get better results.

5. Summary

This research found that viewing data field as a kernel and using topological potential for parameter selecting has better accuracy and shorter run time than four common kernels. Besides, it is very coincident with the Vapnik‘s theory, which guarantees the generalization of learning machines theoretically. This method could be further research with the specific background and applications.

Acknowledgements

The author confirms that this article content has no conflict of interest.

References

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision,* vol. 60, pp. 91-110, 2004.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding,* vol. 110, pp. 346-359, 2008.

[3] K. Lillywhite, D.-J. Lee, B. Tippetts, and J. Archibald, "A feature construction method for general object recognition," *Pattern Recognition,* vol. 46, pp. 3300-3314, 2013.

[4] C. L. Novak and S. A. Shafer, "Anatomy of a color histogram," in *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on*, 1992, pp. 599-605.

[5] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, pp. 273-297, 1995.

[6] . *Color histogram*. Available: http://en.wikipedia.org/wiki/Color_histogram

[7] H. Ling and K. Okada, "Diffusion distance for histogram comparison," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006, pp. 246-253.

[8] Y. Lin, S. Wang, L. Zhao, and D. Wang, "Topology Potential-Based Parameter Selecting for Support Vector Machine," in *Advanced Data Mining and Applications*, ed: Springer, 2014, pp. 513-522.

[9] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *The annals of statistics,* pp. 1171-1220, 2008.

[10] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*: Cambridge university press, 2004.

[11] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American mathematical society,* pp. 337-404, 1950.

[12] A. Aizerman, E. M. Braverman, and L. Rozoner, "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and remote control,* vol. 25, pp. 821-837, 1964.

[13] V. Vapnik, "The nature of statistical learning theory (Information Science and Statistics)," 1999.

[14] V. Vapnik, "Statistical learning theory," ed: Wiley, New York, 1998.

[15] M. Mitchell, "An introduction to genetic algorithms, 1996," *PHI Pvt. Ltd., New Delhi,* 1996.

[16] M. Clerc, "Standard particle swarm optimisation," 2012.

[17] "[Mathworks]Object recognition methods in computer vision."

[18] F.-F. Li, M. Andreetto, M. A. Ranzato, and P. Perona. *Caltech 101*. Available: http://www.vision.caltech.edu/Image_Datasets/Caltech101/