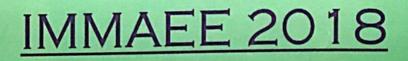
IMMAEE 2018 IOP Conf. Series: Materials Science and Engineering 452 (2018) 011003 doi:10.1088/1757-899X/452/1/011003

Peer review statement

All papers published in this volume of *IOP Conference*. Such Materials Science and Engineering have been peer reviewed through processes administered by the second distribution of th - 244 hu avnart rafare

IOP Publishing



SEPTEMBER 15-16, 2018 - MELBOURNE, AUSTR

2219 3B INITERNATIONAL CONFERENCE OUT

INSULATING MATERIALS, MATERIAL APPLICATION AND ELECTRICAL ENGINEERING

Volume 452, Year 2018



10P Conference Series

IDP Conference Series Materials Science and Engineering

ISSN: 1757-8981

Volume 452, Year 2018

IIOP Institute of Physics.

Research on Test Verification Index system of New Generation Electronic equipment

Jinpeng Yang, Guangyao Lian, Junxian Wang, Ting Ye and Lei Wang

Research on Equivalent Fault Injection Method Based on Dependency Matrix 042183

042182

Sp Keschills J. (APP Split biller on Angentation Space Optice) Control SpectDetection A Statistic in Ma Willing Zha, Williamg, PlongyadLi and Manfang Mai - Yileng Zhu, Yi Huang, Hongyu Li and Yaufang J

	hiders and a second second		n in tes	And	Sec. Sec.	UNIN ALL OF
7 ()	DERVILIMENTAL CONTRACT	later providence in the second				1000 g at 1. C . L
	an ann an san san bhail			, installation (10	E ROMANNE	
					a an	
	ո՞տ II հաստանություն	<u>110</u> 1 u	₩ ₩ ₩₩ ₩ 800, 00% ₩	ши _W III II 801 8008 ₈₈	II	
1111 100 V			λαδία II ~~		∞ n ∞iii ii∞∞∞ u %H 008 H n ‰n ^v ∞ R u	ж м Пахаха 880 ал
€.,¶⊓ ₀			Sinnallinn II allen Minnall I	annalan anna anna ann		Mulaninyan
	. 80000 0 0' 0 "04000" . 088 0 0800 IPoBay x*	"Norther of the state of the st	C Nu auu			
	<mark>x</mark> hu m x 🌘 u huhu	<mark>IIy nyong neg nong 888</mark>	William Syllightill and the	Maxw&avWr IIMIrW a ^{nwr} Wrd	a mu <mark>n</mark> unawa	ma nilitan
uanud 8 x u 8 usu						
Щ. ш. 100 М						
	a <mark>s IIIs")salan si "Willia</mark>		arvar "Musika i	CIII II IIIMII AMBINI Manimiinii	vitestin "sili un ma sun Mille mille de sun Mille mille de sun anno sun anno sun anno sun anno sun anno sun anno	
		8 <mark>0 wallininininininini ya</mark>		WWW ^{oo w} oos ⁿ Wuodhu	**************************************	
0	۲ م ۵۱ ۵۱ <mark>الباس المام م</mark> مس ^ر سا معرفان الم	Woon A Andl Man "WWAnd N 20 A An An An M X A An A	10	MIISAA SE SA MUUUU Maa Se Inaswinaise ^a i W	an a a a an A li ann a an an Ann an a	
s Mau			, 27 ¹⁷⁷ 74 (197 ¹ 911) (1977) 27177 (1977) (1977) 27177 (1977) (1977)	08008800008800008000000000000000000000	isso n numas san mana. Mana	vullansoon on Iv
u nilio US			، ۲۵۵٬۰۰۰ ۵ ۵ ۵ ۵٬۵۵٬۰۰۰ ۱۱ ۱۱۰۰ - ۱۱۰۰ - ۱۱۰۰ - ۱۱۰۰ - ۱۱۰۰ - ۱۱۰۰ - ۱۱۰۰۰		™ _W amaana i Tu muun	1 ⁸⁷⁷ 108 or 8 × 108 11U ¹ uñminanih
um 200 - 10 200 - 10						
u≈"≈ Xull	, ⁸ 00, ⁸⁰ 0, 800, ⁸⁰⁰ , ¹¹ 11 11, 11, 11, 11, 11, 11, 11, 11, 11,	<mark>UIII ~IIII,vontigentijov/118</mark> , v UIII ~III.,wu	n _a sunus ^{sun} nin u <u>s</u> sunus prus 	Manalan Un Munaa wu Manalan Un Munaa wu		°°≊ ∭h∞∭
۳۳۳ ۱۳۰۵ ۱۳۰۵	* "***********************************		אווא _ע יייאא [ָ] _א ואוועזיי אאוון ע א א ע אייייעא אייאאון א	MHAYA "UU" ya Mia shi Ma shi Nasa na walika ha Na sa	*: \$	Man a Man a Masa nn Uta S
	ang group _{an} ang group ang gro					
и 8 0 [®] J			man a construction and a second s International and the second s			
ηΪw			a may and a second s	LuT 01"M XXXII.U0		0 off

¹Nanhang Jincheng College, Nanjing, China
²College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China
³Key Laboratory of Radiometric Calibration and Validation for Environmental Satellites, China Meteorological Administration (LRCVES/CMA), Beijing 100081, China

⁴National Satellite Meteorological Centre, Beijing 100081, China

* Corresponding author e-mail: zxn0215@163.com, ^ajetly@126.com, ^byangcj@cma.gov.cn, ^balan8409@126.com

A machine learning model of handwritten digit recognition based on SVC is established in this paper. Then the influence of sample number, kernel function parameters, penalty coefficients and other parameters on the prediction model is analysed. This results show that training samples have a significant impact on the model. There is an acceptable training number. Different kernel functions have a different effect on the accuracy of the model. The radial basis function is the best in recognition model. The recognition rate increases first with gamma increases, and when gamma increases to a certain value, precision begins to decline.

Image is one of the important means for information to express, store and transmit. Along with the rapid development of the computer technology and digital image processing technology, digital image classification and recognition techniques are required urgently. Image recognition, widely applied to handwriting recognition, face recognition, vehicle license plate recognition and so on, is the use of AI technology to enable computers to recognize information in images. For instance, the methods to recognize handwritten date, account digit and other numeric information are the key technique of image recognition.

Support vector machine (SVM) is a learning method based on statistical learning theories, which proposed by Vapnik. Basing on the principle of structural risk minimization, SVM can improve the generalization ability of the learning machine as much as possible. Even the decision rules obtained from limited training samples can still get small errors for independent test datasets. In recent years, SVM has been widely used in pattern recognition, regression analysis and feature extraction. Vapnik found that different kernel functions had little effect on SVM performance. The key factors affecting SVM performance are kernel function parameters and penalty coefficient. Therefore, the study of kernel function parameters and penalty coefficient is an important field to improve the performance of machine learning. A machine learning model of handwritten numeric information based on SVC is established in this paper. Then the influence of training number, kernel function and penalty coefficients parameters on the

prediction model is analysed, providing reference for improving efficiency and accuracy of machine learning model.

A hyper-plane or set of hyper-planes in a high or infinite dimensional space, which can be used for classification, regression or other tasks, is constructed by a support vector machine. In general, since the larger margin the lower generalization error of the classifier, the hyper-plane, which has the largest distance to the nearest training data points of any class, achieve a good separation.

SVM, including C-SVC, v-SVC, one-class SVM, e-SVR, v-SVR and so on, is a set of supervised learning methods used for classification, regression and border detection. SVCs are classes capable of performing multi-class classification on a dataset.

Given training vectors $x_i \in \mathbb{R}^n$, $i = 1, \dots, l$, in two classes, and a vector $y \in \mathbb{R}^l$, and $y \in \{1, 1\}$, SVC solves the following primal problem:

$$\min_{w,b,} \frac{1}{2} w^{T} w C_{i}^{l} \\
i \\ subject to \\
j \\ 0, i \\ 1, ..., l$$
(2-1)

Its dual is

$$\min \frac{1}{2} \quad {}^{T}Q \quad e^{T}$$
subject to $y^{T} \quad 0$

$$0 \qquad C, i \quad 1, \dots, l$$

$$(2-2)$$

Where e is the vector of all ones, C = 0 is the upper bound, Q is n by n positive semi-definite

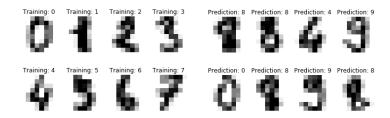
$$k(x,z) \quad (x^T z \quad c)^d \tag{2-5}$$

3) RBF (Radial Basic Function)

$$k(x,z) \exp(||x||^2)$$
 (2-6)

4) Sigmoid kernel function

$$k(x,z) \quad \frac{1}{1 \quad \exp(vx^T \quad c)} \tag{2-7}$$



The training data examples (left) and forecast data examples (right)

3.2.2 Machine learning library

At present, there are many machine learning libraries in Python, of which scikit-learn is the most famous, simple and efficient tools for data mining and data analysis, it can be accessible to everybody, and reusable in various contexts. In scikit-learn, an estimator for classification is a Python object that implements the methods fit(x, y) and predict(T), and some forecast result is show in fig 2.

We are given samples of each of the 10 possible classes (the digits 0 through 9) on which we fit an estimator to be able to predict the classes to which unseen samples belong. The cache_size is set to 500(MB). When training SVM with the kernel function of RBF, C and gamma are considered.

Training samples have a very significant impact on the model. When the number of training samples increases, the running time increases dramatically. The running time of the model also increases with the number of predict, but the influence is not significant as the training samples (fig 3).

As the number of training samples increases, the accuracy of the model will increase significantly. However, when the predicted model reach a certain precision, increasing the number of training samples cannot contribute to the accuracy of the model. When the prediction number are 300, 500 and 800 with the training samples increase from 100 to 600, and the accuracy increases from less than 90% to more than 94%. But the accuracy of the model changes little with training number increasing up to 600(fig 3). So, it is no need to collect unlimited number of samples for training, there is an acceptable training number.

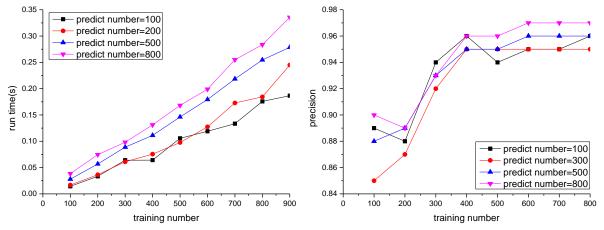


Figure 3. The relation of training number between run time (left) and precision (right)

Different kernel functions have a different effect on the accuracy of the model. The experimental results show that the performance of RBF is the best. The recognition precision of handwritten digits can reach 93% - 100%, and the average recognition precision is 97% (fig 4). The recognition precision

of Linear and Polynomial can also reach 83%-100%, and the average recognition precision is 93% and 94% respectively. However, the generalization performance of Sigmoid kernel function is weak, and the recognition precision vary from 48% to 95%, the precision of digits 2, 3 and 8 is much worse (no more than 55%), and the average precision is only 70%.

RBF has the strongest learning ability, linear and polynomial functions have the stronger learning ability, while sigmoid has the weakest learning ability.

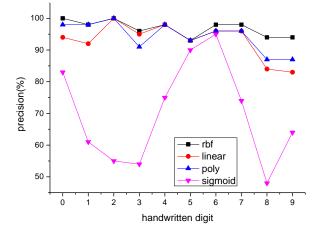


Figure 4. The precisions between different kernel functions

The setting of initial parameters is very important, and the parameters directly affect the generalization ability of SVM. It is necessary to select suitable parameters to train SVC models through tests.

When training an SVM with the RBF kernel, C and gamma is considered. The recognition rate increases continuously with C increases. When C increases to a certain extent, the recognition rate changes little, the maximum accuracy is 0.97. The recognition rate increases first with gamma increases, and when gamma increases to a certain value (about 0.001-0.002), precision begins to decline (fig. 5).

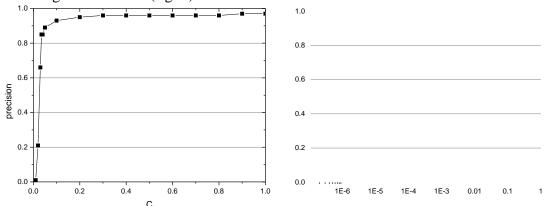


Figure 5. The relationship between C (left),gamma(right) and recognition precision

(1)Training samples have a significant impact on the model. As the number of training samples increases, the accuracy of the model will increase significantly. When the predicted model reach a best precision, increasing the number of training samples cannot improve the accuracy of the model. There is an acceptable training number.

(2)Different kernel functions have a different effect on the accuracy of the model. The experimental results show that the performance of RBF is the best. RBF has the strongest learning ability, linear and polynomial have the stronger learning ability, while sigmoid has the weakest learning ability.

(3) The recognition rate increases continuously with C. When C increases to a certain extent, the recognition rate changes little, the maximum accuracy is 0.97. The recognition rate increases first with gamma increases, and when gamma increases to a certain value, precision begins to decline.

This work was financially supported by xxx fund.

- [1] Vladimir N. Vapnik An Overview of Statistical Learning Theory. IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 10, NO. 5, SEPTEMBER 1999
- [2] Vladimir N.Vapnik. The Nature of Statistical Learning Theory, Springer, NY, (1995)
- [3] Stitson M O et al .Theory of support vector machines, Technical Report CSD-T R-96-17, Depar tment of Computer Science, Roy al Holloway University of London ,(1996)
- [4] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.
- [5] C. Kaynak.