

Image Recognition of Maize Leaf Disease Based on GA-SVM

Zhiyong Zhang^a, Xiaoyang He^{a*}, Xiaohua Sun^b, Limei Guo^c, Junhao Wang^d,
Fushun Wang^d

^a College of Information Science and Technology, Agricultural University of Hebei, Baoding, China

^b Department of Digital Media, Hebei Software Institute, Baoding, China

^c Human resources and social security bureau of Jingxiu area Hebei, Baoding, China

^d College of Information Science and Technology, Agricultural University of Hebei, Baoding, China
hexiaoyang0667@163.com

An improved SVM named genetic algorithm support vector machines (GA-SVM) is discussed in this paper for classifying maize disease. According to the disadvantage of manually determining the parameters in traditional SVM, the genetic algorithms is used to automatically obtain the penalty factor and kernel function. The appropriate parameters are selected by rotational orthogonal method. The extracted eigenvalue is entered to the GA-SVM classification model to improve the classification performance. After the comparisons of different genetic operators and different kernel functions, the results show that the appropriate parameters for genetic algorithms is when $M=50$, $P_c=0.7$ and $P_m=0.05$, the average classification rate is at peak when choosing RBF kernel function. The results also demonstrate that the GA-SVM algorithm achieves a better improvement than SVM.

1. Introduction

Maize is one of the most important food and feed crops with important economic significance in china. With the changes of the global climate, cultivation methods and release of new varieties, the maize diseases are becoming more serious in recent years. In order to realize the intellectual control of the maize disease, the digital image processing technology has already become one of the main technologies which was confirmed (Sanyal P, et al. (2008); Li C M, et al. (2007); Karimi Y, et al. (2006); Huang K Y, et al. (2007); Li C M, et al. (2010)). The different recognition methods are put forward according to the different objects and information. The design and performance of the classifier directly affects the object identified speed and precision which was confirmed (Huerta E B, et al. (2006); Luo Y H, et al. (2011); Liu J H, et al. (2009); Mark J L, et al. (2008); Li B N, et al. (2011)). The common methods of plant disease identification are based on neural network or SVM (support vector machines). BP algorithm is the most popular training algorithm for feed forward neural network learning. But falling into local minimum and slow convergence are its drawbacks. So the SVM method has been applied more widely which was confirmed (Huanga Y, et al. (2010); Lie J, et al. (2006); Sankarana S, et al. (2010)). When traditional support vector machines separate data, the penalty factor and kernel function are manually determined, which is random and blind which was confirmed (Li C, et al. (2008); Patil J K, et al. (2011); Liu H, et al. (2005)). In this paper, GA (genetic algorithm) is applied to automatically obtain the penalty factor and kernel function of SVM. The appropriate parameters are obtained by rotational orthogonal method. The extracted eigenvalue is entered to the GA-SVM classification model to improve the classification performance.

2. Image processing

2.1 Image preprocessing

A series of maize diseases images were collected by digital camera under the sunlight condition. The JPEG format images are converted into BMP format to acquire information. After the gray transformation and the histogram equalization, the combined filter is used on the image to make the segmentation more precision. Figure 1 shows the original images, gray scale images and result of segmentation from top to bottom. Sequencing is: *exserohilum turcicum*, brown spot, gray spot, *curvularia lunata*, round spot and corn rust.

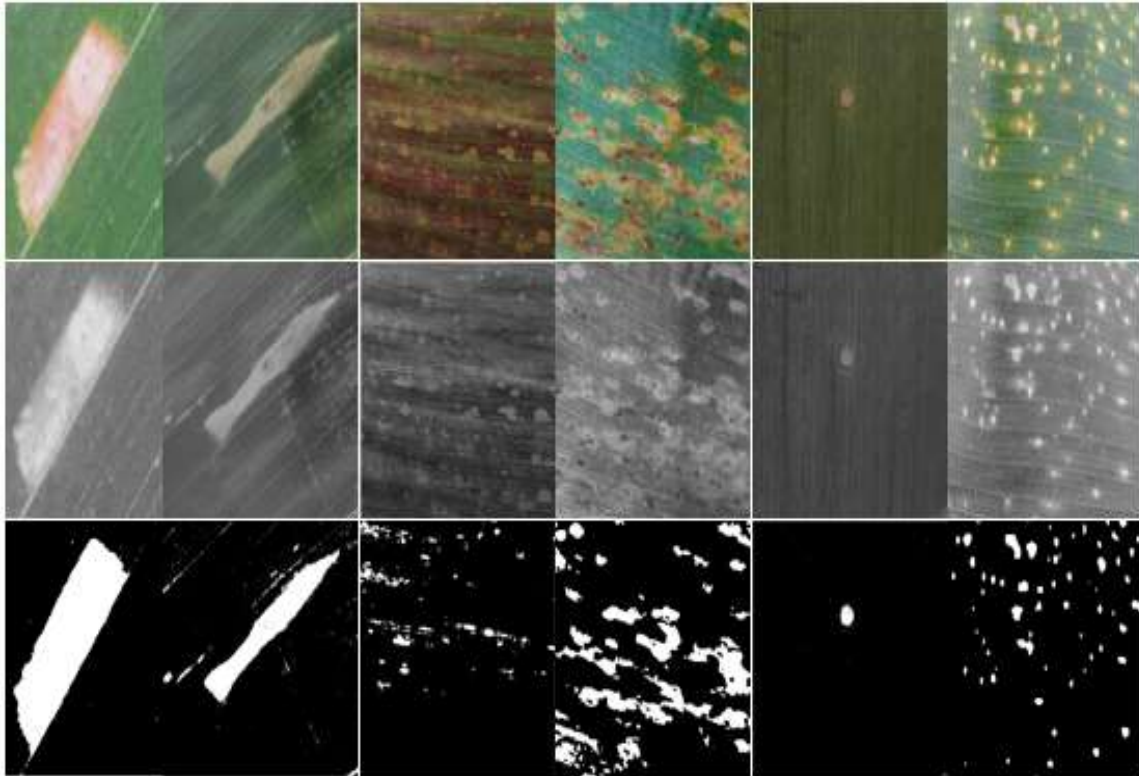


Figure 1: The original Images and preprocessing results

2.2 Feature extraction

Firstly, the image is converted from RGB model to HSI one, and then the mean value and standard deviation of the three components is extracted. Then the binary images are obtained using global threshold segmentation. Then the shape features are extracted, including areas, circumference, circularity, rectangle, complexity and the width and height of minimum enclosing rectangle, as shown in table 1.

Table 1: The extracted features

feature name	expression	feature name	expression
mean value of R	A(R)	mean value of H	A(H)
mean value of G	A(G)	mean value of S	A(S)
mean value of B	A(B)	mean value of I	A(I)
standard deviation of R	STR(R)	standard deviation of H	STR(H)
standard deviation of G	STR(G)	standard deviation of S	STR(S)
standard deviation of B	STR(B)	standard deviation of I	STR(I)
areas	A	circumference	L
circularity	c_i	rectangle	S_q
complexity	c_o	rectangle areas	A_R
width	a	height	b

3. GA-SVM algorithms

The best combinatorial parameter value for GA is obtained through the rotating and perpendicular method. A new method for SVM model selection using genetic algorithm flowchart is presented below.

Step 1: the characters of penalty factor C and kernel σ^2 are encoded with binary arithmetic coding scheme in support vector machine algorithm.

Step 2: a group of individuals are generated at random for population initialization. For increasing the diversity of population, the evenly distributed random numbers are generated primarily. The range of values of the random numbers, a total of M , is allowed in the range 0 to 1. A single gene in any individual is generated by the method:

$$C = C_{\min} + (C_{\max} - C_{\min}) * d \quad (1)$$

$$\sigma^2 = \sigma_{\min}^2 + (\sigma_{\max}^2 - \sigma_{\min}^2) * d, [C, \sigma^2] \quad (2)$$

An initial population is generally generated as $[C, \sigma^2]$, which has the actual size of $M \times 2$.

Step 3: the value of individual genes in the population is adopted as the SVM parameters. The SVM model is trained by training sets.

Step 4: the fitness function of samples x_i is expressed as the formula below.

$$f(x_i) = \text{ano} / \text{pno} \quad (3)$$

The number of data samples classified correctly in testing samples is expressed as ano . The total number of samples is expressed as pno . Comprehensively judge whether the requirements of optimization guidelines is met. If the results show that it can meet the requirements, the best individual is exported and the optimal solution is obtained. The computation is finished. Otherwise turn to Step 5.

Step 5: the parent individuals are selected based on fitness.

The chromosome is chosen according to the selecting probabilities P_s . The selection process is in accordance with the roulette mechanism.

$$P_s = f(x_i) / \sum_{i=1}^M f(x_i) \quad (4)$$

$$f_s = \sum_{i=1}^M f(x_i) \quad (5)$$

Repeat the Step s for selection:

1) Calculate f_s and p_s in t -generation.

2) The random number expressed as a , whose value range is in $[0,1]$ is generated, and $s = a * f_s$ is calculated.

3) The minimum k is found out in $\sum f(x_i) \geq s$, the k th individual can be selected.

4) $t = t + 1$;

5) If the population size reaches M , this procedure ends.

Step 6: crossover and mutation:

The value of crossover probability P_c , mutation probability P_m and evolutionary generation G_m is given. In general, the range of values of crossover probability P_c is in interval $[0.4, 0.9]$; the range of P_m is in interval $[0.001, 0.1]$; the range of G_m is in interval $[100, 500]$. Set several group of P_c , P_m and G_m respectively to do the calculation. The suitable P_c , P_m and G_m are selected by rotational orthogonal method. The parent's chromosomes are randomly selected in population according to crossover probability P_c . The new individuals are generated by the single point crossing method. The chromosomes are selected according to the mutation probability P_m . The new individuals are obtained with uniform mutation method. The chromosomes are listed in order of descending value of the fitness function. The parent chromosomes with the minimum fitness are replaced by the offspring one with the maximum fitness. The update operations are based on the elitist strategy.

Step 7: The new generations are produced through mutation and crossover and go back to (4).

Step 8: the training of genetic algorithm is ended. The optimal parameter of SVM is the value of the gene that has the maximum elimination in current population. The GA-SVM model is built. Calculates its output by the well-trained GA-SVM model and then classification result could be gotten.

The critical conditions for GA-SVM apply one of the rules below.

1) The evolutionary generation exceeds the preset value.

2) The individual maximum fitness value in populations exceeds the preset value.

3) The individual average fitness value in populations exceeds the preset value.

4. Results and analyses

4.1 The comparisons of different genetic operators

A total of 20 feature parameters mentioned in Figure 1 are selected as input vector. SVM with RBF kernel is used in the experiment. The rotational orthogonal experiment is based on different genetic parameters to obtain the GA-SVM optimal model. Then classification of the model for recognition of grape disease is operated. The all three factors of the genetic algorithm are crossover probability P_c , mutation probability P_m and the population scales M . The population scales are separately assigned values of 20, 50, 80 and 100. The crossover probability P_c are separately assigned values of 0.5, 0.7, 0.8 and 0.9. The mutation probability P_m is separately assigned values of 0.001, 0.01, 0.05 and 0.1.

The result of the rotational orthogonal experiment is shown in tab 1. The *exserohilum turcicum*, brown spot, gray spot, *curvularia lunata*, round spot and corn rust are represented by 1 to 6 in table 2.

Table 2: The result of the rotational orthogonal experiment

NO	M	P_c	P_m	classification rate/%						average/%
				1	2	3	4	5	6	
1	20	0.5	0.001	88.68	85.79	90.35	88.78	85.89	90.45	88.32
2	20	0.7	0.01	88.65	87.21	90.50	88.76	87.32	90.61	88.84
3	20	0.8	0.05	88.79	86.09	88.06	88.90	86.20	88.17	87.70
4	20	0.9	0.10	88.26	85.38	85.13	88.37	85.49	85.24	86.31
5	50	0.5	0.10	88.50	86.51	86.57	88.61	86.62	86.68	87.25
6	50	0.7	0.05	91.25	88.72	88.54	91.36	88.82	88.65	89.56
7	50	0.8	0.01	88.98	88.68	90.64	89.09	88.78	90.74	89.49
8	50	0.9	0.001	88.69	87.57	91.10	88.79	87.68	91.21	89.17
9	80	0.5	0.05	89.26	87.34	88.76	89.37	87.45	88.86	88.51
10	80	0.7	0.10	89.12	86.21	86.63	89.23	86.32	86.74	87.38
11	80	0.8	0.001	89.32	87.60	91.17	89.43	87.71	91.28	89.42
12	80	0.9	0.01	88.68	88.14	90.71	88.78	88.25	90.81	89.23
13	100	0.5	0.01	88.65	88.93	90.76	88.76	89.03	90.87	89.50
14	100	0.7	0.001	90.02	89.36	91.21	90.13	89.46	91.32	90.25
15	100	0.8	0.10	89.36	86.32	86.69	89.46	86.43	86.80	87.51
16	100	0.9	0.05	89.18	86.42	88.94	89.29	86.52	89.04	88.23

The trial data demonstrates that the average classification rate are at peak of 90.25% when $M=100$, $P_c=0.7$ and $P_m=0.001$. The second peak classification rate is when $M=50$, $P_c=0.7$ and $P_m=0.05$. the classification rate is 89.56%. Though it is a little lower than that of the maximum, its population size is only half of the former. So the appropriate parameters for genetic algorithms is when $M=50$, $P_c=0.7$ and $P_m=0.05$.

4.2 The classification comparisons of different kernel functions

The test samples are the same. Set $M=50$, $P_c=0.7$ and $P_m=0.05$. The linear kernel, polynomial kernel, RBF kernel and sigmoid kernel are selected respectively to compare. The result of the different kernel functions is shown in table 3.

Table 3: Classification results of the different kernel functions

kernel function	classification rate/%						average/%
	1	2	3	4	5	6	
linear	88.21	69.74	90.48	92.62	73.23	95.00	84.88
polynomial	86.80	86.70	90.48	91.14	91.04	95.00	90.19
RBF	92.59	89.62	89.45	97.22	94.10	93.92	92.82
sigmoid	49.01	49.01	49.01	51.46	51.46	51.46	50.24

The trial data demonstrate that the average classification rate is at peak when choosing RBF kernel function. The average classification rate can reach 92.82%, which has far exceeded the sigmoid kernel function.

4.3 The comparisons of different classification method

The test samples are also the same. The classifications are by two methods of GA-SVM and SVM separately. The result of the different methods is shown in table4.

Table 4: Classification results of the different algorithms

disease	classification rate/%		C	σ^2
	GA-SVM	SVM		
1	92.59	90.09	251.4283	0.0002
2	89.62	69.74	256.6615	0.0414
3	89.45	85.59	191.9561	0.4590
4	96.61	92.90	248.8886	0.0002
5	88.72	69.63	254.0689	0.0410
6	88.55	81.48	190.0171	0.4543

The trial data demonstrates that the GA-SVM algorithm achieves an improvement. Especially in the brown spot and corn rust identification, the GA-SVM has increased almost 20 percent.

5. Conclusions

The image analysis-based research on maize disease diagnosis recognition technology has a great significance to prevent the occurrence of disasters. In this paper, the GA-SVM algorithm can automatically retrieve kernel parameter values. By contrast, the SVM algorithm must be manually assigned parameter values after experiment. So the new algorithm is reliable and efficient. The research contents of this paper provides a theoretical basis for the identification and diagnosis of maize leaf disease, it also can provide technical support for the maize leaf disease intelligence, automation, control, and safety in production.

Acknowledgments

This work is supported by rural informatization engineering technology research center of Hebei province and 2015 annual Science and Engineering Foundation of Hebei Agricultural University, China. (Grant No. LG201506, LG20150602).

References

- Huang K.Y. 2007. Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features. *Computers and Electronics in Agriculture*, vol. 57, no. 1, pp. 3-11, doi: 10.1016/j.compag.2007.01.015

- Huang Y.B., Lan Y.B., Thomson S.J., et al. 2010. Development of soft computing and applications in agricultural and biological engineering. *Computers and Electronics in Agriculture* vol. 71, pp. 107-127, doi: 10.1016/j.compag.2010.01.001
- Huerta E.B., Beatrice D., Hao J.K. 2006. A hybrid GA/SVM approach for gene selection and classification of micro array data. *Lecture Notes in computer Science*, vol. 39, no. 7, pp. 34-44
- Karimi Y., Prasher S.O., Patel R.M, Kim S.H. 2006. Application of support vector machine technology for weed and nitrogen stress detection in corn. *Computers and electronics in agriculture*, vol. 51, pp. 99-109, doi: 10.1016/j.compag.2005.12.001
- Li B.N., Chui C.K., Chang S., et al. 2011. Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation. *Computers in Biology and Medicine*, vol. 41, no. 1, pp. 1-10
- Li C., Kao C., Gore J.C., Ding Z. 2008. Minimization of region-scalable fitting energy for image segmentation. *IEEE Transaction on Image Processing*, vol. 17, no. 10, pp. 1940-1949
- Li C.M., Kao C.Y., Gore J., et al. 2007. Implicit Active Contours Driven by Local Binary Fitting Energy. In, pp. *Computer Vision and Pattern Recognition, CVPR '07. IEEE Conference on. Minneapolis, Minnesota, USA.* IEEE, pp. 1-7
- Li C.M., Xu C.Y., Senior Member, et al. 2010. Distance regularized level set evolution and its application to image segmentation. *IEEE Transaction on Image Processing*, vol. 19, no. 12, pp. 3243-3254
- Lie J., Lysaker M., Tai X.C. 2006. A binary level set model and some applications to mumford-shah image segmentation. *IEEE Transaction on Image Processing*, vol. 15, no. 5, pp. 1171-1181, doi: 10.1109/TIP.2005.863956
- Liu H., Yu L. 2005. Toward integrating feature selection algorithms for classification and clustering [J]. *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491-502
- Liu J.H. 2009. Improvement and the basic theory of particle swarm optimization. Changsha: Central South University, 22-27
- Luo Y.H., Fu C.Y. 2011. Midfrequency-based real-time blind image restoration via independent component analysis and genetic algorithms. *Opt. Eng.*, vol. 50, no. 4, 047004, doi: 10.1117/1.3567072
- Mark J.L., Michael K N, Cheung Y M, et al. 2008. Agglomerative fuzzy K-means clustering algorithm with selection of number of cluster. *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 11, pp. 1519-1534
- Patil J.K., Kumar R. 2011. Advances in image processing for detection of plant diseases. *Journal of Advanced Bioinformatics Applications and Research*, vol. 2, no. 2, pp. 135-141
- Sankarana S., Mishraa A., Ehsani R., et al. 2010. A review of advanced techniques for detecting plant diseases, *Computers and Electronics in Agriculture*, vol. 72, 1-13, doi: 10.1016/j.compag.2010.02.007
- Sanyal P., Patel S.C. 2008. Pattern recognition method to detect two diseases in rice plants. *Applied engineering in agriculture*, vol. 56, no. 6, pp. 319-325, doi: 10.1179/174313108X319397