

Handwriting Classification Based on Support Vector Machine with Cross Validation

Anith Adibah Hasseim, Rubita Sudirman, Puspa Inayat Khalid

Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia

Email: rubita@fke.utm.my

Received 2013

ABSTRACT

Support vector machine (SVM) has been successfully applied for classification in this paper. This paper discussed the basic principle of the SVM at first, and then SVM classifier with polynomial kernel and the Gaussian radial basis function kernel are chosen to determine pupils who have difficulties in writing. The 10-fold cross-validation method for training and validating is introduced. The aim of this paper is to compare the performance of support vector machine with RBF and polynomial kernel used for classifying pupils with or without handwriting difficulties. Experimental results showed that the performance of SVM with RBF kernel is better than the one with polynomial kernel.

Keywords: Support Vector Machine; Handwriting Difficulties; Cross-Validation

1. Introduction

The field of handwriting has been of interest from a variety of aspects; its entity, indications and aesthetic. In the beginning, the development of handwriting and the factors that affect handwriting performance were investigated [1,2], but later whole words were addressed. Most of the systems reported in the literature until today involved screening measures in identifying pupils who are at risk of handwriting difficulties and also addressed the absence of an appropriate tool for monitoring beginning handwriting development. More importantly, automated handwriting analysis has been given more attention in the hunt for quantitative features and key indicators in monitoring beginning handwriting skill development. Such automated handwriting analysis include recognizing the writer (e.g. [3]), the text written (e.g. [4]), movement and procedure (e.g. [5,6]), or even semantic content of the text (e.g. [7]). More or less each of these issues can, and have been investigated either offline or online related to the available data.

Up to sixty percent of children's typical school day is allocated to fine motor activities, with writing being the predominant task during these time periods [8]. These tasks all require the foundational skill of basic handwriting proficiency to allow teachers to accurately assess students' understanding and comprehension of instructional material. If students do not possess basic handwriting proficiency, it can limit their ability to successfully complete a majority of classroom tasks. In addition, it has also been suggested that students with handwriting problems need to focus more attention on the physical

process of writing, thus limiting use of higher order cognitive skills, planning and generation of content [9]. Thus, handwriting proficiency is an important foundation upon which success with later writing tasks depends. Due to the number of every day school tasks which involve writing, unsuccessful mastery of handwriting skill can negatively influence later success in school.

1.1. Support Vector Machine

Support Vector Machine (SVM) is a new classification technique based on the statistical learning theory proposed by Vapnik in 1995 [10]. It can successfully solve over-fitting, local optimal problem and is especially suitable for small-sample and high-dimensional nonlinear case. Besides, it already showed good results in the medical diagnostics, optical character recognition, electric load forecasting and other fields.

Kernel Function

In general, a radial basis function is one of the most popular kernel and reasonable first choice. The reason why is, this kernel nonlinearly

Given the linearly separability sample set (x_i, y_i) where $i = 1, \dots, n$. If taking the simplest case; 2 class classification, then $x \in R^n$, $y \in \{+1, -1\}$ is the classes number. The commonly form of the linear decision function is:

$$f(x) = w \cdot x + b \quad (1)$$

Sometimes linear classifiers are not complex enough; therefore SVM maps the data into a higher dimensional space, unlike the linear kernel which can handle the case

when the relation between class labels and attributes is nonlinear [11]. Formally, pre-process the data with:

$$x \rightarrow \varphi(x) \quad (2)$$

and then learn the map from $\varphi(x)$ to y :

$$f(x) = w \cdot \varphi(x) + b \quad (3)$$

However, the dimensionality of $\varphi(x)$ can be very large, making w hard to represent explicitly in memory, and hard to solve. The Representer theorem (Kimeldorf & Wahba, 1971) shows that (for SVMs as a special case):

$$w = \sum_{i=1}^m \alpha_i \varphi(x_i) \quad (4)$$

for some variables α . Instead of optimizing w directly we can thus optimize α . The decision rule is now:

$$f(x) = \sum_{i=1}^m \alpha_i \varphi(x_i) \cdot \varphi(x) + b \quad (5)$$

If the dot product ($x \cdot x_j$) is replaced by the kernel function $K(x, x')$, the optimal decision function is as follows:

$$f(x) = \sum_{i=1}^m \alpha_i K(x_i, x) + b \quad (6)$$

In this project, 2 kinds of common kernel function are used. The first one is Gaussian radial basis function (RBF):

$$K(x_i, x) = \exp\left(-\frac{x - x_i^2}{g^2}\right) \quad (7)$$

and the other one is polynomial kernel:

$$K(x_i, x) = [(x \cdot x_i) + 1]^d \quad (8)$$

Classical techniques utilizing radial basis functions employ some method of determining a subset of centre. Typically a method of clustering is first employed to select a subset of centre. An attractive feature of the SVM is that this selection is implicit, with each support vectors contributing one local Gaussian function, center at that data point.

1.2. Cross Validation (CV)

Currently, cross-validation has been widely used for estimating the performance of neural networks and other applications such as support vector machine and k-nearest neighbor. Cross-validation is a statistical method of evaluating and comparing learning algorithms. The basic idea of cross-validation is splitting the data, which is consists of dividing the available training data into two sets. The first set is used to train the network, while the other is used to evaluate the performance of the trained network. In typical cross-validation, the training and validation sets must cross-over in successive rounds such that each data point has a chance of being validated

against. The basic form of cross-validation is k-fold cross-validation. Other forms of cross-validation are special cases of k-fold cross-validation or involve repeated rounds of k-fold cross-validation.

Advantages of this method are as follows: 1) Average classification accuracies of k SVM classifiers are used to evaluate the SVM classifier parameters performance which can improve the generalization ability of the SVM classifier with the optimized parameters; 2) k -fold cross-validation method can ensure all the sample data be involved in the SVM classifier training and validation, it can make full use of the limited sample data; 3) no matter how the data gets divided, every data point is used as a test set exactly once, and gets to be in a training set $k-1$ times. The disadvantage of this method is that the training algorithm has to be rerun from scratch k times, which means it takes k times as much computation to make an evaluation.

2. Methodology

The data was obtained from Khalid *et al* in [13]. The data is composed of 120 samples which contain 2 features (that is The standard deviation of pen pressure when drawing RU, p-value < 0.0001 and z-value = minus 4.319 and Ratio of time taken to draw HR and HL, p-value < 0.0001 and z-value = minus 5.205.) and two group of writers (that is below average printers (test group) and above average printers (control group)).

Firstly, the data is portioned into k equally sizes segments or folds. In this project, we used 10-fold cross validation ($k = 10$) as it is the most common used for data mining and machine learning. As shown in **Figure 1**, the darker section of the data are used for training while the remaining data; lighter sections are used for validate the model. This process is repeated 10 times until all sections have been validated.

Model Parameter Selection

Two models; SVM of polynomial kernel function and RBF kernel are chose in looking for performance comparison. Performance of the SVM depends on the choice of parameters. The optimal selection of these parameters is a nontrivial issue. According to study, the important of RBF kernel is need to find parameter C and g . SVM of polynomial kernel function chooses different parameter C and d . The penalty factor C , is used to improve generalized capability when C is increasing while g and d are the adjustable parameter of study machine in the experiment and they are used to adjust experienced error value. The parameter slightly influences classification result when a smaller amount of training samples are used [12].

After training SVM, the best value C and g can be used to classify children with handwriting problems. For

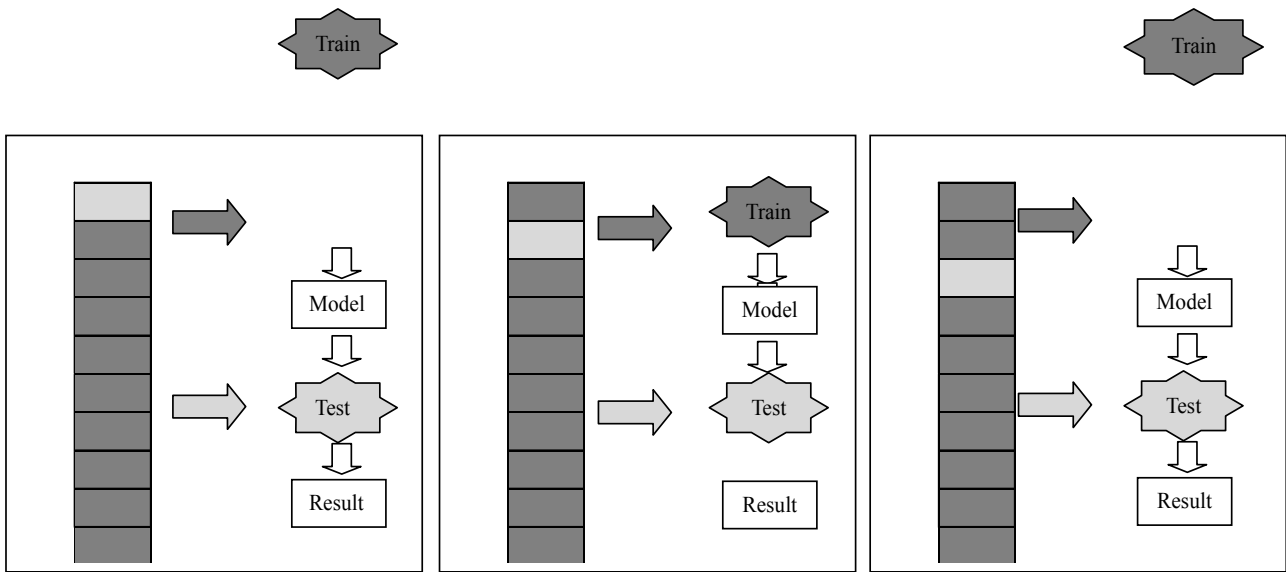


Figure 1. Procedure of 10-fold cross validation.

the SVM with polynomial kernel, there are two parameters: C and d . The SVM with RBF kernel has also two parameters: g and C . In order to know different performance each parameter produces to outputs, we select three values for each parameter just like choosing the number of hidden nodes in the neural networks.

3. Results and Discussion

Table 1 and Table 2 present the recognition results using the SVM with polynomial kernel, RBF kernel respectively. The classification was considered correct if the output from the model was similar to the one that had been judged by the teachers (using Handwriting Proficiency Screening Questionnaire (HPSQ)). In this paper, we used the classification error (rejection of genuine category) as the metric.

According to Table 1, as can be seen the percentage of correct prediction of feature 1 is in decreasing when the variation g varies from 0.01 to 0.1. While it is in reverse direction when the variation g varies from 0.1 to 1. The results confirmed that the best value of the variation g near 0.1. When the coefficient of penalty C is increased, the accuracy of prediction is in decreasing. Different from feature 1, feature 2 is seen to be decreasing in percentage of correct prediction when g varies from 0.01 to 1 and when C increases in the value.

In the other hand, the result from Table 2 shows in different from Table 1. It is clear that, when the variation g increases in the value from 0.01 to 1, both percentage of corrects prediction for feature 1 and feature trends to decrease. While when the variation d varies from 3 to 10, the accuracy of prediction is increasing. This exhibits SVM good generalization performance.

Table 1. Accuracy of Prediction based on SVM with RBF Kernel.

Feature	g/C	Accuracy of Predictions (%)		
		0.01	0.1	1
Feature 1	1	83.33	91.67	83.33
	10	93.33	91.67	91.67
	100	91.67	91.67	91.67
Feature 2	1	91.67	92.80	91.67
	10	91.67	91.67	83.33
	100	91.67	86.67	83.33

Table 2. Accuracy of Prediction based on SVM with Polynomial Kernel.

Feature	g/d	Accuracy of Predictions (%)		
		0.01	0.1	1
Feature 1	3	86.67	86.67	91.67
	5	83.33	86.67	83.33
	10	66.67	71.67	83.33
Feature 2	3	86.67	86.67	86.67
	5	83.33	83.33	86.67
	10	61.67	71.67	86.67

The results reported here have shown that the performance of SVM with RBF kernel is better than SVM with polynomial kernel. We use SVM (RBF kernel) with

changing C and g to simulate and to classify children with and without handwriting problem based on drawing tasks.

4. Conclusions

SVM RBF and polynomial have been used in this study to select those who are at risk of handwriting difficulty due to the improper use of graphic rules. Cross-validation method is adopted to choose parameter in order to gain preferable classificatory result. In this paper, we have testified that the performance of SVM with RBF kernel is better than the one with polynomial kernel. Experiment simulative results indicate: average accuracy of classificatory testing based on SVM RBF algorithm reaches more than 93%. The data is apparently high compared with SVM polynomial algorithm.

5. Acknowledgements

This work was supported by the Malaysia Ministry of Higher Education and Universiti Teknologi Malaysia under Vote Q.J130000.2623.09J28.

REFERENCES

- [1] V. Berninger, A. Cartwright, C. Yates, H. L. Swanson and R. Abbott, "Developmental Skills Related to Writing and Reading Acquisition in the Intermediate Grades: Shared and Unique Variance," *Reading and Writing: An Interdisciplinary Journal*, Vol. 6, 1994, 161-196. [doi:10.1007/BF01026911](https://doi.org/10.1007/BF01026911)
- [2] S. Graham, V. W. Berninger, N. Weintraub and W. Schafer, "Development of Handwriting Speed and Legibility in Grades 1-9," *Journal of Educational Research*, Vol. 92, 1997, pp. 42-52. [doi:10.1080/00220679809597574](https://doi.org/10.1080/00220679809597574)
- [3] Z. Yong, T. Tan and Y. Wang, "Biometric Personal Identification Based on Handwriting," *Pattern Recognition, Proceedings. 15th International Conference on*, Vol. 2, 2000, pp. 797-800.
- [4] L. M. Lorigo and V. Govindaraju, "Offline Arabic Handwriting Recognition: A Survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 28, pp. 712-724, 2006. [doi:10.1109/TPAMI.2006.102](https://doi.org/10.1109/TPAMI.2006.102)
- [5] H. Ishida, et al., "A Hilbert Warping Method for Handwriting Gesture Recognition," *Pattern Recogn.*, Vol. 43, 2010, pp. 2799-2806. [doi:10.1016/j.patcog.2010.02.021](https://doi.org/10.1016/j.patcog.2010.02.021)
- [6] H. Bezine, A. D. Alimi and N. Sherkat, "Generation and Analysis of Handwriting Script with the Beta-Elliptic Model," *Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition*, 2004, pp. 515-520. [doi:10.1109/IWFHR.2004.45](https://doi.org/10.1109/IWFHR.2004.45)
- [7] S. Srihari, J. Collins, R. Srihari, H. Srinivasan, S. Shetty and J. B. Griffler, "Automatic Scoring of Short Handwritten Essays in Reading Comprehension Tests," *Artificial Intelligence*, Vol. 172, 2008, pp. 300-324. [doi:10.1016/j.artint.2007.06.005](https://doi.org/10.1016/j.artint.2007.06.005)
- [8] K. McHale and S. A. Cermak, "Fine Motor Activities in Elementary School: Preliminary Findings and Provisional Implications for Children with Fine Motor Problems," *American Journal of Occupational Therapy*, Vol. 46, No.10, 1992, pp. 898-903. [doi:10.5014/ajot.46.10.898](https://doi.org/10.5014/ajot.46.10.898)
- [9] V. Berninger, "Coordinating Transcription and Text Generation in Working Memory During Composing: Automatic and Constructive Processes," *Learning Disability Quarterly*, Vol. 22, 1999, pp. 99-112. [doi:10.2307/1511269](https://doi.org/10.2307/1511269)
- [10] V. N. Vapnik, "The Nature of Statistical Learning Theory," Springer, New York, 1995. [doi:10.1007/978-1-4757-2440-0](https://doi.org/10.1007/978-1-4757-2440-0)
- [11] M. Hong, G. Yanchun, W. Yujie and L. Xiaoying, "Study on Classification Method Based on Support Vector Machine," *2009 First International Workshop on Education Technology and Computer Science, Wuhan, China*, 7-8 March 2009, pp.369-373.
- [12] S. S. Keerthi and C. J. Lin, "Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel," *Neural Computation*, Vol. 15, No. 7, 2003, 1667-1689. [doi:10.1162/089976603321891855](https://doi.org/10.1162/089976603321891855)
- [13] S. S. Keerthi and C. J. Lin, "Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel," *Neural Computation*, Vol. 15, No. 7, 2003, pp. 1667-1689. [doi:10.1162/089976603321891855](https://doi.org/10.1162/089976603321891855)