CHAPTER 4

4. WRITER IDENTIFICATION MODEL THROUGH SVM WITH EXISTING KERNELS

This chapter describes the implementation of a writer identification models using Support Vector Machine (SVM). The standard SVM algorithm builds a binary classifier by constructing a hyper plane which separates to classes of data. SVM automatically identifies a subset of informative points called support vectors and uses them to represent the separating hyper plane. Three independent experiments were carried out here based on three datasets and are explained with the flow of event like building the model, performance and evaluation in this chapter. The results of performance evaluation with various measures of the classifiers are also presented and the findings are summarized.

4.1 MODEL I - WRITER IDENTIFICATION MODEL USING SUPPORT VECTOR MACHINE

In modeling Tamil Handwriting Writer Identification (THWI) [25], the essential tasks such as corpus preparation, preprocessing, feature extraction, building the model are carried out. In this work, Support Vector Machine based classifiers are developed using linear, polynomial and RBF kernels for multi class classification. Efficient writer identification models are built by tuning the regularization, degree, gamma parameters. The predictive performance of the classifiers is evaluated using various metrics like predictive accuracy, precision, recall, F-measure, time taken and the results are analyzed.

Building the Model

The training data set with 26000 instances are used for training SVM. $\text{SVM}^{\text{light}}$ is used for implementation. The training datasets and the test datasets are converted into the format required by $\text{SVM}^{\text{light}}$ [55]. Profile of the datasets is shown in Table 4.1.

Table 4.1 Profile of the Datasets

Dataset	Character	Word	Paragraph
Total Number of Instances	30000	30000	30000
Number of Instances in Training Dataset	24000	24000	24000
Number of Instances in Testing Dataset	6000	6000	6000
Number of Features	26	26	422
Number of Class Labels	1-300	1-300	1-300

Various kinds of kernels such as linear, polynomial and RBF kernel are used in SVM training with different parameter settings for d, gamma and C as regularization parameter. The parameters d and gamma are associated with polynomial kernel and RBF kernel respectively. The values of the regularization parameter C is assigned between 0.5 and 50 for linear kernel. For polynomial and RBF kernels the value for C is assigned as 0.5, 1 and 5, d is assigned from 1 to 4 and g is taken from 0.5 to 5 respectively. It is found that the regularization parameter reaches a stable state for the value C = 5. Three independent writer identification models have been built. The sample screenshots of the learning and classification process are shown in Fig. 4.1 and Fig. 4.2.

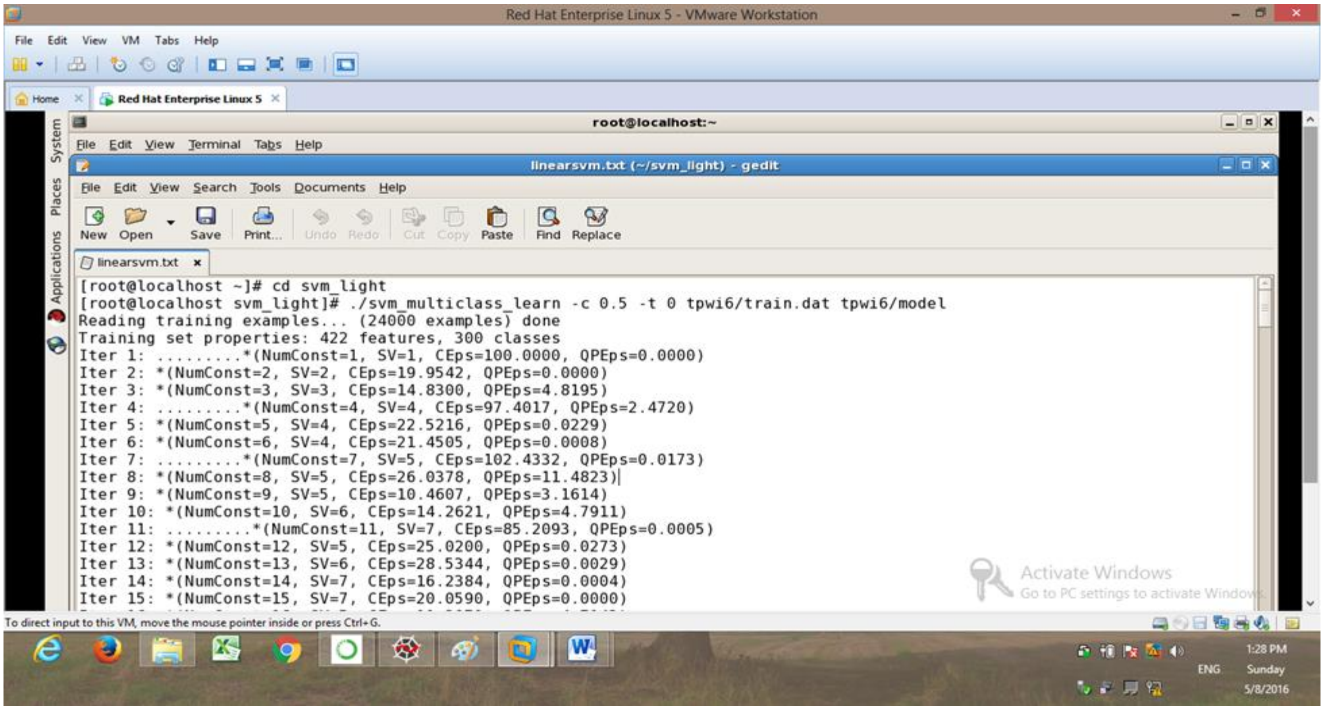

Fig. 4.1 Learning SVM with Linear Kernel

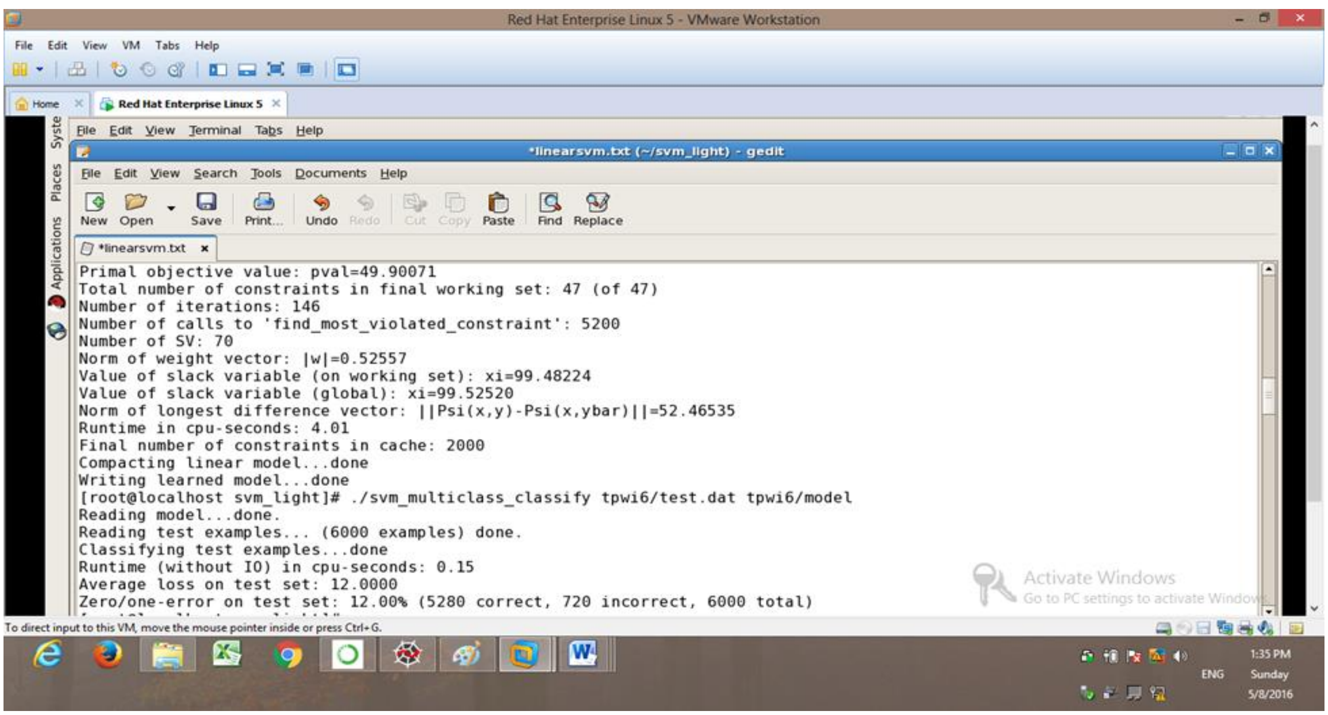

Fig. 4.2 Classification using Linear Kernel SVM Model

Performance Evaluation

The respective test sets are used to evaluate the performance of the writer identification models. Various evaluation metrics like precision, recall, F-measure and accuracy have been considered and the results are obtained. The results of SVM based writer identification models built using TWINC (character) dataset with linear, polynomial and RBF kernels is shown in Table 4.2 - Table 4.4 and the comparative performance is shown in Table 4.5 and illustrated in Fig. 4.3.

Table 4.2 Results of SVM with Linear Kernel (Character)

Parameters	**C=25**	**C=15**	**C=30**
No. of CCI	4236	4032	4110
No. of ICCI	1764	1968	1890
No. of SV	121	138	129
Accuracy (%)	70.6	67.2	68.5
Time Taken (in secs)	0.02	0.01	0.02
Precision	0.732	0.698	0.711
Recall	0.962	0.891	0.922
F-measure	0.821	0.788	0.803

Table 4.3 Results of SVM with RBF Kernel (Character)

Parameters	C=0.5			C=1			C=5		
G	**1.5**	**3.5**	**4**	**1.5**	**3.5**	**4**	**1.5**	**3.5**	**4**
No. of CCI	5436	5340	5418	5400	5340	5412	5436	5400	5346
No. of ICCI	564	660	582	600	660	588	564	600	654
No. of SV	115	123	116	117	123	116	115	117	122
Accuracy (%)	90.6	89	90.3	90	89	90.2	90.6	90	89.1
Time Taken (in secs)	0.03	0.02	0.03	0.02	0.02	0.03	0.03	0.02	0.02
Precision	0.91	0.88	0.90	0.91	0.88	0.92	0.91	0.91	0.79
Recall	0.88	0.90	0.89	0.83	0.90	0.92	0.88	0.83	0.95
F-measure	0.89	0.89	0.78	0.87	0.89	0.92	0.89	0.87	0.85

Table 4.4 Results of SVM with Polynomial Kernel (Character)

Parameters	C=0.5		C=1		C=5	
D	**3**	**4**	**3**	**4**	**3**	**4**
No. of CCI	4590	5352	5118	4752	5340	5352
No. of ICCI	1410	648	882	1248	660	648
No. of SV	139	122	132	134	126	122
Accuracy (%)	76.5	89.2	85.3	79.2	89	89.2
Time Taken (in secs)	0.4	0.62	0.8	0.4	0.61	0.8
Precision	0.77	0.87	0.79	0.79	0.85	0.90
Recall	0.95	0.91	0.95	0.84	0.89	0.90
F-measure	0.85	0.89	0.86	0.81	0.75	0.85

Table 4.5 Consolidated Results of all Three SVM Models (Character)

Kernels	Linear	RBF	Polynomial
No. of CCI	4236	5436	5352
No. of ICCI	1764	564	648
No. of SV	121	115	122
Accuracy (%)	70.6	90.6	89.2
Time Taken (in secs)	0.02	0.03	0.62
Precision	0.732	0.91	0.87
Recall	0.962	0.88	0.91
F-measure	0.821	0.89	0.89

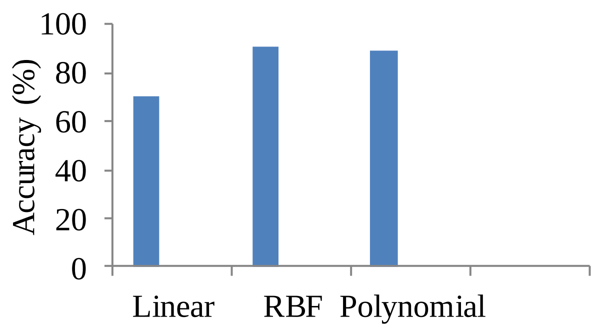

Fig. 4.3 Comparative Results of Accuracy (Character)

From the above comparative analysis it is observed that the RBF kernel based prediction model (90.6%) shows high accuracy than the polynomial (89.2%) and linear kernel (70.6%) SVMs and average time taken to build the model is high in SVM with polynomial kernel (0.62) than the other models. Other kernels such as linear and RBF takes only 0.02 and 0.03 respectively. As far as machine learning is concerned, accuracy plays a major role in evaluating the performance of the predictive models than time taken. Hence it is concluded that SVM with RBF kernel (90.6%) based writer recognition model out performs well.

The results of second experiment of SVM based writer identification models built using TWINW (word) dataset with linear, polynomial and RBF kernels is shown in Table 4.6 – Table 4.8 and the comparative performance is shown in Table 4.9 and illustrated in Fig. 4.4.

Table 4.6 Results of SVM with Linear Kernel (Word)

Parameters	**C=5**	**C=12**	**C=24**
No. of CCI	4500	4320	4392
No. of ICCI	1500	1680	1608
No. of SV	143	168	152
Accuracy (%)	75	72	73.2
Time Taken (in secs)	0.03	0.02	0.03
Precision	0.705	0.733	0.744
Recall	0.749	0.968	0.971
F-measure	0.726	0.834	0.842

Table 4.7 Results of SVM with RBF Kernel (Word)

Parameters	C=0.5			C=1			C=5		
G	**1.5**	**3.5**	**4**	**1.5**	**3.5**	**4**	**1.5**	**3.5**	**4**
No. of CCI	5628	5520	5538	5628	5520	5526	5628	5520	5526
No. of ICCI	372	480	462	372	480	474	372	480	474
No. of SV	128	136	133	128	136	136	128	136	133
Accuracy (%)	93.8	92	92.3	93.8	92	92.1	93.8	92	92.1
Time Taken (in secs)	0.05	0.06	0.05	0.05	0.06	0.05	0.05	0.06	0.05
Precision	0.92	0.926	0.916	0.92	0.926	0.906	0.92	0.926	0.906
Recall	0.99	0.916	0.920	0.99	0.916	0.724	0.99	0.916	0.724
F-measure	0.96	0.920	0.29	0.96	0.920	0.804	0.96	0.920	0.804

Table 4.8 Results of SVM with Polynomial Kernel (Word)

Parameters	C=0.5		C=1		C=5	
D	**3**	**4**	**3**	**4**	**3**	**4**
No. of CCI	4710	5472	5040	5472	5460	5472
No. of ICCI	1290	528	960	528	540	528
No. of SV	175	141	152	135	141	149
Accuracy (%)	78.5	91.2	84	91.2	91	91.2
Time Taken (in secs)	0.44	0.69	0.9	0.44	0.69	0.9
Precision	0.755	0.919	0.75	0.919	0.892	0.919
Recall	0.722	0.879	0.90	0.879	0.948	0.879
F-measure	0.738	0.895	0.81	0.895	0.918	0.895

Table 4.9 Consolidated Results of all Three SVM Models (Word)

Kernels	Linear	RBF	Polynomial
No. of CCI	4500	5628	5472
No. of ICCI	1500	372	528
No. of SV	143	128	141
Accuracy (%)	75	93.8	91.2
Time Taken (in secs)	0.03	0.05	0.69
Precision	0.705	0.92	0.919
Recall	0.749	0.99	0.879
F-measure	0.726	0.96	0.895

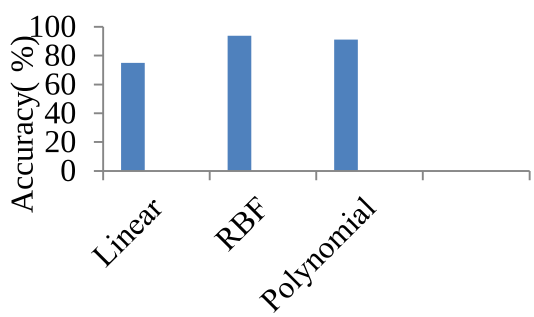

Fig. 4.4 Comparative Results of Accuracy (Word)

From the above comparative analysis it is observed that the RBF kernel based prediction model (93.8%) shows high accuracy than the polynomial (91.2%) and linear kernel (75%) SVMs and average time taken to build the model is high in SVM with polynomial kernel (0.69) than the other models. Other kernels like linear and RBF takes only 0.03 and 0.05 secs respectively. Hence it is concluded that SVM with RBF kernel (93.8%) based writer recognition model out performs well.

The results of next experiment of SVM based writer identification models built using TWINP (paragraph) dataset with linear, polynomial and RBF kernels is shown in Table 4.10 – Table 4.12 and the comparative performance is shown in Table 4.13 and illustrated in Fig. 4.5.

Table 4.10 Results of SVM with Linear Kernel (Paragraph)

Parameters	C=0.5	C=1	C=5
No. of CCI	5160	5160	5280
No. of ICCI	840	840	720
No. of SV	165	360	462
Accuracy (%)	86%	86%	88%
Time Taken (in secs)	3.07	3.70	4.01
Precision	0.796	0.792	0.942
Recall	0.956	0.958	0.989
F-measure	0.869	0.866	0.964

Table 4.11 Results of SVM with RBF Kernel (Paragraph)

Parameters	C=0.5			C=1			C=5		
G	0.5	2	3	0.5	2	3	0.5	2	3
No. of CCI	5688	5646	5574	5688	5520	5574	5688	5520	5688
No. of ICCI	312	354	426	312	480	426	312	480	312
No. of SV	162	298	397	162	178	251	129	184	111
Accuracy (%)	94.8	94.1	92.9	94.8	92	92.9	94.8	92	94.8
Time Taken (in secs)	12.45	13.52	13.51	12.52	11.66	12.78	12.47	12.62	12.71
Precision	0.962	0.960	0.930	0.951	0.928	0.922	0.954	0.924	0.966
Recall	0.722	0.868	0.653	0.609	0.762	0.577	0.649	0.725	0.646
F-measure	0.825	0.912	0.767	0.743	0.837	0.710	0.772	0.813	0.775

Table 4.12 Results of SVM with Polynomial Kernel (Paragraph)

Parameters	C=0.5		C=1		C=5	
D	1	2	1	2	1	2
No. of CCI	4800	5616	5160	5616	5406	5616
No. of ICCI	1200	384	840	384	594	384
No. of SV	458	115	336	110	380	123
Accuracy (%)	80	93.6	86	93.6	90.1	93.6
Time Taken (in secs)	2536.49	8654.58	2838.27	8635.12	2624.16	7256.89
Precision	0.807	0.93	0.796	0.93	0.90	0.93
Recall	0.874	0.99	0.958	0.99	0.973	0.99
F-measure	0.838	0.95	0.869	0.95	0.935	0.95

Table 4.13 Consolidated Results of all Three SVM Models (Paragraph)

Kernels	Linear	RBF	Polynomial
No. of CCI	5280	5688	5616
No. of ICCI	720	312	840
No. of SV	462	162	115
Accuracy (%)	88	94.8	93.6
Precision	0.942	0.962	0.93
Recall	0.989	0.722	0.99
F-measure	0.964	0.825	0.95
Time Taken (in secs)	4.01	12.71	8654.58

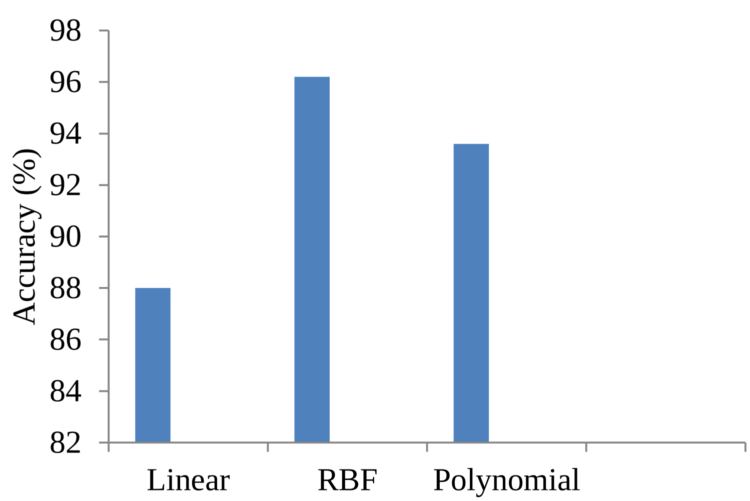

Fig. 4.5 Comparative Results of Accuracy (Paragraph)

From the above comparative analysis it is observed that the RBF kernel based prediction model (94.8%) shows high accuracy than the polynomial (93.6%) and linear kernel (88%) SVMs and average time taken to build the model is high in SVM with polynomial kernel (8654.58) than the other models. Other kernels, linear and RBF takes only 4.01 and 12.71 secs respectively. Hence it is concluded that SVM with RBF kernel (94.8%) based writer recognition model out performs well.

Findings

It is proved that increase in number of instances in the training datasets helps to build efficient models. It is also observed that RBF kernel based models built using all three datasets are more appropriate and reasonable for writer identification. It is found that prediction accuracy is high in case of SVM models with RBF kernel than other kernels. The novel idea of combining local and global features designed for building the classifier in paragraph text images is found even more decisive in identifying the writing pattern than the existing models. The comparative performance analysis shows that handwriting with more words and sentences i.e., paragraph text offers more contributive features and hence the corresponding model yields more accuracy in distinguishing the individuals than character and word text.

4.2 SUMMARY

This chapter demonstrates the modeling of writer identification as the problem of learning multiclass classification that suits to identify writer effectively. It describes the implementation of Tamil writer identification using SVM for three levels of handwritings character, word and paragraph text images. The experiments carried out in SVM^{light} are described and the results are presented in tables and charts. The findings of this work are also summarized in this chapter.

Remarks

1. A paper titled “Discovering Tamil Writer Identity Using Global and Local Features of Offline Handwritten Text” has been published in the Journal of International Review on Computers and Software (IRECOS), Italy, Vol. 8 No 9, Page 2080 – 2087, 2013.
2. A paper titled “Prediction of Writer Using Tamil Handwritten Document Image Based on Pooled Features”, has been published in International Journal of World Academy of Science, Engineering and Technology, International science index, Vol. 9, No. 6, Page 1481 – 1487, 2015.

3. A paper titled "Detection of a Person Using Descriptive Features of Tamil Handwriting and Pattern Learning", has been published in International Journal of Applied Engineering Research, **(Scopus Indexed)**, ISSN 0973-4562, Vol 10(21), 2015, pp 41902-41909.