Biometric Recognition using Feature Selection and Combination

Ajay Kumar[1, 2], David Zhang[2]

[1]Department of Electrical Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, 110016, India
[2]Depertment of Computing, Hong Kong Polytechnic University, Hung Hom, Hong Kong
ajaykr@ieee.org, csdzhang@comp.polyu.edu.hk

Abstract. Most of the prior work in biometric literature has only emphasized on the issue of feature extraction and classification. However, the critical issue of examining the usefulness of extracted biometric features has been largely ignored. Feature evaluation/selection helps to identify and remove much of the irrelevant and redundant features. The small dimension of feature set reduces the hypothesis space, which is critical for the success of online implementation in personal recognition. This paper focuses on the issue of feature subset selection and its effectiveness in a typical bimodal biometric system. The feature level fusion has not received adequate attention in the literature and therefore the performance improvement in feature level fusion using feature subset selection is also investigated. Our experimental results demonstrate that while majority of biometric features are useful in predicting the subjects identity, only a small subset of these features are necessary in practice for building an accurate model for identification. The comparison and combination of features extracted from hand images is evaluated on the diverse classification schemes; naive Bayes (normal, estimated, multinomial), decision trees (C4.5, LMT), k-NN, SVM, and FFN.

1 Introduction

Feature evaluation is critical while designing a biometric based recognition system under the framework of supervised learning. The existing research in biometrics has not made any attempt to evaluate the usefulness of the features that have been proposed in the literature [1]-[4]. Feature subset selection helps to identify and remove much of the irrelevant and redundant features. The small dimension of feature set reduces the hypothesis space, which is critical for the success of online implementation in personal recognition. Furthermore, researchers have shown [5]-[8] that the irrelevant and redundant training features adversely effects the classifier performance. Table 1 summarizes the effect of redundant training information on some common machine learning algorithms. These observations provide us the motivation to perform the experiments to evaluate the advantages of the feature subset selection and combination of some common biometric modalities.

Most of the prior work in the biometric fusion literature has examined the fusion of modalities at score and decision level. However it is generally believed that a fusion scheme applied as early as possible in the recognition system is more effective [9]. Therefore, the fusion at feature level typically results in a better improvement than at decision level. This is because the feature representation conveys the richest information as compared to the matching scores or abstract labels. Furthermore, the feature level fusion has not received adequate attention in the biometrics. Therefore we have also investigated the performance improvement using feature level fusion in the context of feature subset selection.

Table 1. Effect of redundant biometric features

Machine Learning Algorithm	***Effect of Irrelevant Features***	***References***	***Typical Example***
Nearest Neighbor	Training complexity grows exponentially	[5],[6]	[1],[2]
Naive Bayes	Invalidation of assumption that features are independent in a given class	[7]	[3]
Decision Trees	Overfitting of training data, Large tree complexity	[8]	[4]

2 Feature Evaluation and Selection

Feature selection is used to identify the useful features and remove the redundant information. The usage of small size feature vector results in reduced computational complexity which is critical for online personal recognition. The selection of effective features may also result in increased accuracy. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) have been traditionally used to reduce the large dimension of feature vectors. However, PCA or LDA *transforms* the feature vectors to reduced dimension rather than actually selecting a subset. Several feature subset selection algorithm have been proposed in the literature. The feature subset evaluation and selection algorithm consists of three basic modules as shown in Figure 1; feature subset-generation and –evaluation and the stopping criteria. Let N be the total number of potential biometric features in the biometric training dataset. The exhaustive search through the total number of 2^N candidate feature subsets in infeasible even with moderate N. Therefore various search strategies (*e.g*. starting point, search direction, *etc*.) have been studied in the literature. The goodness of each of the candidate feature subset is evaluated with a feature evaluation criterion. The goodness index of current feature subset is compared against those from the previous best feature subset and replaced if this index from current feature subset is better than those from the previous best feature subset. There are two commonly used feature evaluation criterion; wrapper-based and filter-based. The wrapper is one of the most commonly used algorithms, which evaluates and selects feature subset by repeated use of a particular classification algorithm. However, it is highly time consuming and prohibitive when the dimension of feature vectors is large (such as those from palmprints evaluated in this work). Therefore we employed filter-based algorithm for the feature evaluation which is detailed in next section. The feature selection process usually stops with a suitable stopping criteria; *e.g.* predefined number of features, iterations or goodness index, addition or deletion of features does not increase goodness index, *etc*.. In this work we used Correlation based Feature Selection (CFS) algorithm which has been shown [10] to be quite effective in feature subset selection. The CFS is a classifier independent algorithm and its usefulness is illustrated from our experimental results.

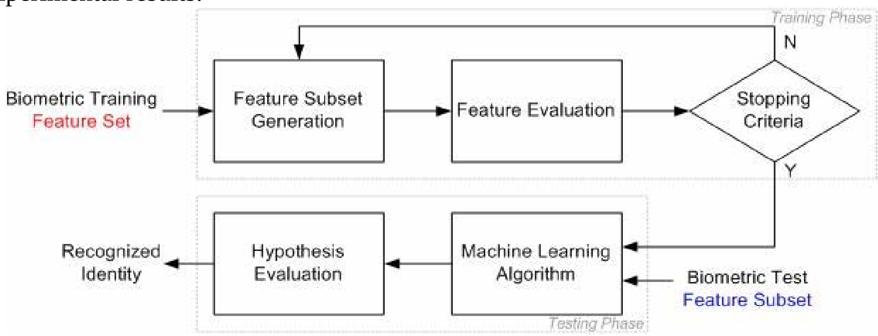

Fig. 1. Building a biometric recognition system using feature subset selection

3 Correlation based Feature Selection Algorithm

The CFS algorithm uses a correlation based objective function to evaluate the usefulness of the features. The objective function $J_{cfs}(\lambda)$, also know as Pearson's correlation coefficient, is based on the heuristic that a good feature subset will have high correlation with the class label but will remain uncorrelated among themselves.

$$J_{cfs}(\lambda) = \frac{\lambda \psi_{cr}}{\sqrt{\lambda + \lambda(\lambda - 1)\psi_{rr}}} \tag{1}$$

Above equation illustrates the merit of λ features subset where ψ_{cr} is the average feature to class correlation and ψ_{rr} is average feature to feature correlation within the class. The CFS based feature selection algorithm uses $J_{cfs}(\lambda)$ to search the feature subsets using the best first search [11]. The best search starts with the evaluation of all the individual features considering them as a separate subset. The feature subset with highest objective function is retained. The new feature subset is further expanded by adding all possible combinations of new single features and all the resulting combinations are evaluated in the similar manner. If the addition of a new feature does not show any improvement then the search process returns to next unexpended subset and repeats in the similar manner. The search is aborted if the addition of new features does not show any improvement in the last 5 consecutive expanded combinations. This stopping criteria is the same (default) used in *MLC*++ machine learning library [12] but for the wrapper feature selector. The expanded feature subset before the termination of search process is assumed to be the best feature subset. However there may be some locally predictive features in the unselected feature set which may be useful in some classification schemes [11]. Therefore the average correlation of every unselected feature with its corresponding class is also examined.

If this correlation is higher than the highest average correlation between any of the already selected features, then this feature is included in the list of best feature subset. We examined the usefulness of CFS scheme by evaluating recognition accuracy and the size of best feature subset with those from original feature set.

4 Classification Schemes

The personal recognition will use the feature vectors from the training images to train or learn the classification algorithm. In this work, we investigated a number of classification algorithms to evaluate the benefits of feature subset selection. These algorithms are quite popular and well known in pattern recognition literature. However, with few notable exceptions, their usefulness for hand recognition is yet to be evaluated. The simplified version of Bayes rule, known as *naive Bayes*, which assumes that the feature vectors within a class are independent, was firstly evaluated. The *naive Bayes* has shown to work well with real data samples and it traditionally makes the assumption that the feature values are normally distributed. However, this assumption may be violated in some domains and our experiments were not restricted to the normality assumption. The distribution of features was also estimated using nonparametric kernel density estimation [14] and employed in the *naive Bayes* classifier. The *multinomial* model has been shown to outperform [15] other alternative models on the real data and was therefore also investigated for the performance.

The *k*-Nearest Neighbor (*k-NN*) classifier employed minimum Euclidean distance between the query feature vector and all the prototype training data. The Support Vector Machine (*SVM*) classifier employed polynomial kernel as it gave us the best results. The execution speed of multi-layer Feed-Forward Neural Network (*FFN*) is among the fastest of all models currently in use. Therefore this network may be the only practical choice for online personal recognition. A linear activation function was selected for the last layer of *FFN* while the sigmoid activation function was employed for other layers. The training weights were updated by using resilient backpropagation, which achieves faster convergence and conserves memory [19].

The decision tree algorithms use training data to build a logical tree and have been proved popular in practice. The *C4.5* [20] is the most used algorithm and uses entropy criteria to select the most informative features for the branching during the training stage. The feature that gives the most information is selected to be at the root of the tree. Another extension of *C4.5*, also known as logistic model tree (*LMT*), uses a combination of tree structure and logistic regression model to build the decision tree. The different logistic regression functions at tree leaves are built using *LogitBoost* algorithm [22]. The construction of *LMT* is detailed in [21] and was evaluated in the experiments as it achieved much higher accuracy than *C4.5*.

5 Experiments

In order to examine the goals of our experiments the biometric image database from 100 subjects was employed. The dataset consisted of 1000 hand images, 10 images per subject, which were obtained from digital camera using unconstrained peg-free setup in indoor environment. These hand images were collected during two sessions with an average interval of three months, as the focus of experiments was to investigate the performance of biometric modalities instead of their stability with time. The volunteers were in the age group of 16-55 years but not too cooperative and were not paid for the data collection. During the image acquisition, the users were only required to make sure that (i) their fingers do not touch each other and (ii) most of their hand (back side) touches the imaging table. The automated segmentation of hand-shape and palmprint image was achieved as detailed in reference [23].

5.1 Palmprint and Hand-Shape Features

Each of the 300 × 300 pixels segmented palmprint images were further divided into 24 × 24 pixel blocks with an overlapping of 6 pixels. The palmprint feature vector of dimension 1 × 144 is extracted from the standard deviation of significant discrete cosine coefficients in each of the image sub-blocks as detailed in [24]. The hand-shape features of dimension 1 × 23 features were also extracted from the hand-shape image; perimeter (f_1), solidity (f_2), extent (f_3), eccentricity (f_4), *x-y* position of centroid relative to shape boundary $(f_5 - f_6)$, convex area (f_7), 4 finger length $(f_8 - f_{11})$, 8 finger width $(f_{12} - f_{19})$, palm width (f_{20}), palm length (f_{21}), hand area (f_{22}), and hand length (f_{23}). Further details on these seven new shape features $(f_1 - f_7)$ can be seen in [24]-[25]. The signature analysis on the hand-shape boundary image is used to extract the image reference points, *i.e*. four fingertips, four interfinger points and hand-base. We employed five image samples from every user collected during the first session for training and the rest for testing. In order to allow fair selection and combination of features, same training and testing splits are used to generate the results.

5.2 Classifier Parameters

The parameters of *SVM* and *FFN* employed in the experiments were empirically selected. The *SVM* using polynomial kernel gave much better results than those from radial basis function. Therefore to conserve the space only results from polynomial kernel are reported. The *SVM* training was achieved with *C-SVM*, a commonly used *SVM* classification algorithm [26]. The training parameter γ and ε were empirically fixed at 1 and 0.001 respectively. Similarly the number of input nodes in *FFN* were also empirically selected for the best performance; 100 (80) for palmprint, 50 (50) for hand-shape and 125 (75) for the combined feature set. The entries in the brackets represent the numbers when corresponding feature subset is employed for the performance evaluation. The *FFN* neuron weights were updated using resilient backpropagation algorithm and the training was aborted if the maximum number of training steps reached to 1000. The *C4.5* decision tree was pruned with a confidence factor of 0.25. The splitting criteria for *LMT* was the same as the one used for *C4.5*, *i.e*. information gain. The minimum number of feature vectors at which a node can be considered for splitting was fixed to 15.

6 Results

The experimental results for the palmprint recognition are summarized in Table 2. This table also shows the performance of corresponding classifier with and without the feature subset selection. The evaluation of 144 palmprint features from the training set, using the CFS algorithm described in section 3, has revealed 75 redundant and irrelevant features. This suggests that the feature selection has been aggressively pursued in the palmprint domain. The performance of 69 relevant palmprint features, or feature subset, is also illustrated in Table 2. It can be seen from this table that the kernel density *estimation* has managed to improve *naive Bayes* performance but the performance improvement is significant when *multinomial* event model is employed. The best performance for palmprint recognition is achieved with *SVM* classifiers when the second order polynomial kernel is used. However, the achieved performance of nearest neighbor classifier suggest that it may be preferred in some applications as it is inherently simple and does not require training phase. The performance of *FFN* is better than *naive Bayes* but quite similar to that from *SVM* or *k-NN*. The performance of decision tree *C4.5* has been worst and this may be due to the large number of features that make the repeated portioning of data difficult. However the performance of *LMT* is also promising and similar to that of *k-NN*. The average tree size for the decision tree build using 144(69) features for *LMT* and *C4.5* was 16 (12) and 285 (281) respectively. This is not surprising as *LMT* algorithm has shown [21] to be often more accurate than *C4.5* and always resulting in a tree of small size than those from *C4.5*.

Table 2. Comparative performance evaluation for the palmprint recognition

	Naive Bayes			KNN	SVM			FFN	Decision Tree	
	Normal	*Estimated*	*Multinomial*		*d*=1	*d*=2	*d*=3		*C4.5*	*LMT*
144 Feature	69.4	74.4	91.8	94.4	95.2	95.8	95.8	95	50.6	94.6
69 features	68.8	75.4	92.8	95	94.4	95.6	95.5	94.8	50.2	93.8

Table 3. Comparative performance evaluation for the hand-shape recognition

	Naive Bayes			KNN	SVM			FFN	Decision Tree	
	Normal	*Estimated*	*Multinomial*		*d*=1	*d*=2	*d*=3		*C4.5*	*LMT*
23 Feature	73.9	79	29	84.6	89	88.4	88.6	86.4	67	89.6
15 Features	78.6	80	51.8	84.2	87	85.4	83.2	85	66.6	87.8

One of the important conclusions from the Table 2 is that the usage of feature selection has effectively reduced the number of features by 52.08% while improving or maintaining similar performance in most cases. *This suggest that while majority of palmprint features are useful in predicting the subjects identity, only a small subset of these features are necessary in practice for building an accurate model for identification*.

Table 3 summarizes the experimental results for the hand-shape identification. The evaluation of 23 hand-shape features from the training data has selected 15 most informative features; $f_1, f_7, f_8 - f_{11}, f_{14}, f_{16} - f_{19}, f_{20} - f_{23}$. The decision tree using *LMT* achieved the best performance while those from the *multinomial naive Bayes* is the worst. The usage of *multinomial* event model in *naive Bayes* has resulted in significant performance improvement from the palmprint features (Table 2) while those from hand-shape features has been degraded (Table 3). This can be attributed to the inappropriate estimation of the term probabilities resulting from the small size hand-shape feature vectors. The average size of decision tree build using 23 (15) features using *LMT* and *C4.5* was 81 (69) and 251 (255) respectively.

Table 4. Comparative performance evaluation for the combined features

	Naive Bayes			KNN	SVM			FFN	Decision Tree	
	Normal	*Estimated*	*Multinomial*		*d*=1	*d*=2	*d*=3		*C4.5*	*LMT*
167 Feature	77	80	80.2	97.4	97.6	98	97.8	97.2	68.4	96
75 Features	78.2	78.6	94.2	97.8	97.8	97.8	98	96.8	68.2	96.4

The experimental results for the combined hand-shape and palmprint features are shown in Table 4. The CFS algorithm selected 75 features subset from the combined list of 167 features. The combined feature subset had 13 hand-shape features, *i.e.* $f_3, f_7, f_8, f_{10} - f_{14}, f_{17} - f_{18}, f_{20} - f_{23}$, and 62 palmprint features. It may be noted that the reduced feature subset obtained from the combined feature set is not the addition or sum of reduced feature subset individually obtained from palmprint and hand-shape feature set. *This suggests that only a certain combination of features, rather than the combination of individual feature subset carrying the discriminatory information, is useful in the feature level fusion*. The new hand-shape features selected in the individual and combined feature subsets, *i.e.* f_3, f_3, f_7, justify their usefulness. However, other new examined hand-shape features, *i.e.* f_2, f_4, f_5, f_6, could not establish their significance. As shown in Table 4, the *SVM* classifier achieved the best performance, which is closely followed by *k-NN*. It can be noticed that the combination of hand-shape and palmprint features has been useful in improving the performance for all the classifiers except for the case from *naive Bayes* classifier. The performances of combined features from the *multinomial naive Bayes* classifier using feature subset selection suggest that the *multinomial* event model is most sensitive to irrelevant and redundant features. The size of decision tree build using 167 (75) features using *LMT* and *C4.5* was 16 (12) and 285 (231) respectively.

It is prudent to examine how the performance of various classifiers that are adversely effected by the irrelevant and redundant features. The performance improvement of these classifiers with the availability of more features, using a fixed number of training samples, is investigated. In this set of experiments, all the available features from the training samples were ranked in the order of their merit using CFS objective function (1). The feature vectors in the test data set were also ranked in the same order of ranking generated from the training data. The performance of these classifiers starting from first 10 features was computed and the next 10 features were added at every successive iterations. The number of input nodes for *FFN* classifier was empirically fixed to 75, irrespective of number of features. Figure 2(a) shows the performance variation for *k-NN*, *SVM* and *FFN* classifiers with the increase in number of features. The *SVM* classifier does not show any appreciable increase in the performance with the addition of irrelevant features (say beyond 75) and its performance is generally the best of all the classifiers evaluated in this paper.

It is interesting to note that the feature selection strategy has been able to find 20 (10) best features that give ≈ 96% (89%) accuracy using SVM classifier; This 20(10) feature subset consists of 15(6) palmprint and 5 (4) hand-shape features.

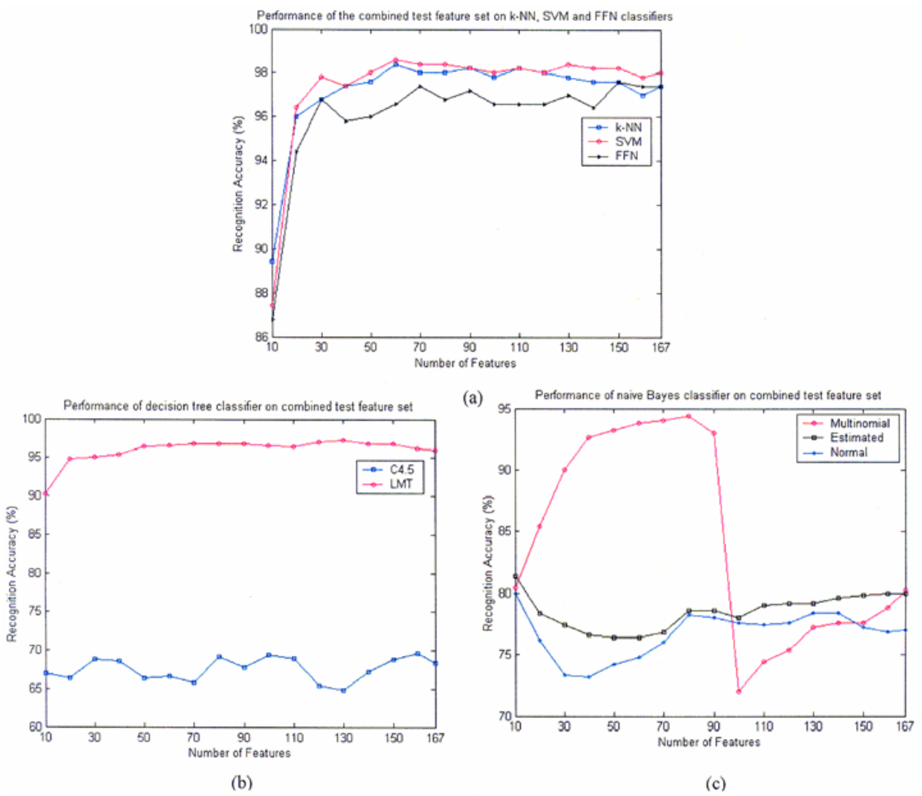

Fig. 2. The performance analysis of classifiers with the number of features; *k-NN*, *SVM* and *FFN* in (a), decision trees in (b), and naïve Bayes in (c)

The performance of *LMT* classifier in Figure 2(b) show an initial increase in performance with the increase in informative features but the performance stabilizes with the addition of non-informative and redundant features (beyond 70-75). Thus the performance of *LMT* suggests that it is insensitive to the redundant and irrelevant features and this is due to the fact that the *LMT* is built using the stagewise fitting process to construct the logistic regression models which select only relevant features from the training data. The *C4.5* decision tree continues to maintain worse performance and the feature selection strategy do not have any appreciable effect on the performance. Figure 2(c) shows the results for the performance of *naive Bayes* classifier. The performance estimates of *naive Bayes multinomial* classifier shows a tendency of exponential increase with small number of features before abrupt decrease in performance. The performance of *naive Bayes* with nonparametric kernel *estimation* is marginally better than those from with *normal* assumption but is quite poor.

7 Conclusions

The evaluation and selection of useful biometric features can improve the accuracy and reduce the complexity of classifier. This can also have high impact on user convenience and economics in the acquisition of online biometric data. It is not possible to locate the relevant features from the real biometrics data in advance and therefore the performance of feature selection strategy must be measured indirectly. The best way to do this is to compare the classifier performance with and without feature subset selection. The effectiveness of feature subset selection and combination was evaluated on the diverse classification schemes; probabilistic classifier (naive Bayes), decision tree classifier (*C4.5, LMT*), and instance based classifier (*k*-NN) and learned classifiers (*SVM, FFN*).

Experimental studies in this paper further suggest that while a majority of features extracted from the hand images are useful in biometric recognition, only a small subset of these features are actually needed in practice for building an accurate model for subject recognition. This is important as the prior studies in biometric literature have not focused on the issue of feature subset selection. The usage of small size feature vectors results in reduced computational complexity, which is critical for online personal recognition. The analysis of experimental results in Table 2-4 indicate that the correlation-based feature subset selection is capable of effectively selecting the relevant palmprint and hand-shape features. We are currently working to examine the issue of feature subset selection for other biometric modalities.

References

1. C. Liu and H. Wechsler, "Independent component analysis of Gabor features for face recognition," IEEE Trans. Neural Networks, vol. 14, pp. 919-928, Jul. 2003.
2. D. Zhang, W.K. Kong, J. You, and M. Wong, "On-line palmprint identification," IEEE Trans. Patt. Anal. Machine Intell., vol. 25, pp. 1041-1050, Sep. 2003.
3. R. Sanchez-Reillo, C. Sanchez-Avila, A. Gonzales-Marcos, "Biometric identification through hand geometry measurements," IEEE Trans. Patt. Anal. Machine Intell., vol. 22, pp. 1168-1171, Oct. 2000.
4. A. Ross and A. K. Jain, "Information fusion in Biometrics", Pattern Recognition Lett., vol. 24, pp. 2115-2125, Sep. 2003.
5. P. Langley and S. Sage, "Scaling to domains with irrelevant features," Computational Learning Theory and Neural Learning Systems, R. Greiner (Editor), vol. 4, MIT Press 1994.
6. D. W. Aha, D. Kibler, and M. K. Albert, "Instance based learning algorithms," Machine Learning, vol. 6, pp. 37-66, 1991.
7. P. Langley and S. Sage, "Induction of selective Bayesian classifiers," Proc. 10th Intl. Conf. Uncertainty in Artificial Intelligence, Seattle, W. A., Morgan Kaufmann, 1994.
8. J. R. Quinlan, C4.5: Programs for machine learning, Morgan Kaufmann, Los Altos, California, 1993.

9. A. K. Jain, A. Ross and S. Prabhakar, " An Introduction to Biometric Recognition", IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, Vol. 14, No. 1, pp. 4-20, January 2004.
10. M. A. Hall and L. A. Smith, "Practical feature subset selection for machine learning," Proc. 21st Australian Computer Science Conference, Springer-Verlag, pp. 181-191, 1998.
11. M. A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," Proc. 7th Intl. Conf. Machine Learning, Stanford University, CA. Morgan Kaufmann Publishers., 2000.
12. R. Kohavi, G. John, R. Long, D. Manley, and K. Pfleger, MLC++: A machine learning library in C++, available on http://www.sgi.com/tech/mlc/docs.html
13. C. Oden, A. Ercil, and B. Buke, "Combining implicit polynomials and geometric features for hand recognition," Pattern Recognition Letters, vol. 24, pp. 2145-2152, 2003.
14. G. H. John and P. Langley, "Estimating continious distribution in Bayesian classifiers," Proc. 11th Conf. on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers, San Mateo, 1995.
15. S. Eyheramendy, D. Lewis, and D. Madigan, "On the naive Bayes model for text classification," to appear in Artificial Intelligence & Statistics, 2003.
16. A. McCallum and K. Nigam, "A comparison of event model for naive Bayes Text Classification," Proc. AAAI-98 Workshop on Learning for Text Categorization, 1998.
17. D. W. Aha, D. Kibler, and K. Albert, "Instance based learning algorithms," Machine Learning, vol. 6, pp. 37-66, 1991.
18. V. Vapnik, Statistical Learning Theory, Wiley & Sons, Inc., New York, 1998.
19. M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The PROP algorithm," Proc. Intl. Conf. Neural Networks, vol. 1, pp. 586-591, Apr. 1993
20. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufman, 1993.
21. N. Landwehr, M. Hall, and E. Frank, "Logistic Model Trees," Proc 14th European Conf. Machine Learning, Cavtat-Dubrovnik, Croatia, Springer-Verlag, Vol. *2837,* pp 241-252, 2003.
22. J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting" The Annals of Statistic, 38(2), 337-374, 2000.
23. A. Kumar, D. C. M. Wong, H. Shen, and A. K. Jain, "Personal verification using palmprint and hand geometry biometric," Proc. AVBPA, pp. 668-675, June 2003.
24. A. Kumar and D. Zhang, "Integrating shape and texture for hand verification," Proc. ICIG 2004, pp. 326-329, Hong Kong, Dec. 2004.
25. John C. Russ, The Image Processing Handbook, 3rd ed., CRC Press, Boca Eaton, Florida, 1999.
26. N. Cristianini and J. S-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, 2001.