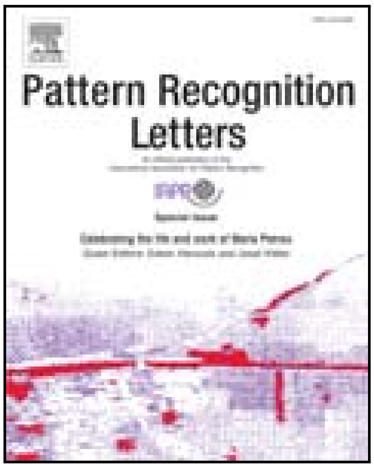
Accepted Manuscript



A Multi-Feature Selection Approach for Gender Identification of Handwriting based on Kernel Mutual Information

Ning Bi, Ching Y Suen, Nicola Nobile, Jun Tan

PII: S0167-8655(18)30171-5
DOI: 10.1016/j.patrec.2018.05.005
Reference: PATREC 7171

To appear in: *Pattern Recognition Letters*

Received date: 8 November 2017
Revised date: 22 April 2018
Accepted date: 3 May 2018

Please cite this article as: Ning Bi, Ching Y Suen, Nicola Nobile, Jun Tan, A Multi-Feature Selection Approach for Gender Identification of Handwriting based on Kernel Mutual Information, *Pattern Recognition Letters* (2018), doi: 10.1016/j.patrec.2018.05.005

Research Highlights (Required)

To create your highlights, please type the highlights against each `\item` command.

It should be short collection of bullet points that convey the core findings of the article. It should include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point.)

- propose a kernel mutual information approach to select features;
- predict the gender of the writers from their handwriting samples;
- handwritten features extracted from different methods, include geometrical and transformed features;
-
-

Pattern Recognition Letters
journal homepage: www.elsevier.com

A Multi-Feature Selection Approach for Gender Identification of Handwriting based on Kernel Mutual Information

Ning Bi[a], Ching Y Suen[b], Nicola Nobile[b], Jun Tan[a,**]

[a]Sun Yat-sen University, Guangzhou, 510275, China
[b]Concordia University, Montreal, QC, H3G 1M8,Canada

ABSTRACT

This paper presents a new flexible approach to predict the gender of the writers from their handwriting samples. Handwriting features like slant, curvature, line separation, chain code, character shapes, and more, can be extracted from different methods. Therefore, the multi-feature sets are irrelevant and redundant. The conflict of the features exists in the sets, which affects the accuracy of classification and the computing cost. This paper proposes an approach, named Kernel Mutual Information (KMI), that focuses on feature selection. The KMI approach can decrease redundancies and conflicts. In addition, it extracts an optimal subset of features from the writing samples produced by male and female writers. To ensure that KMI can apply the various features, this paper describes the handwriting segmentation and handwritten text recognition technology used. The classification is carried out using a Support Vector Machine (SVM) on two databases. The first database comes from the ICDAR 2013 competition on gender prediction, which provides the samples in both Arabic and English. The other database contains the Registration-Document-Form (RDF) database in Chinese. The proposed and compared methods were evaluated on both databases. Results from the methods highlight the importance of feature selection for gender prediction from handwriting.

1. Introduction

The prediction of gender from handwriting samples has recently gained much attention. It has many applications, for example, in forensic analysis where the investigation of crimes usually involves a handwriting examination. Forensic examiners may have to identify the handwritings of individuals. Gender prediction can help investigators to focus on a certain group of suspects. The relationship between handwriting and personality remains disputed and is not completely evaluated by scientific metrics (Tett and Palmer, 1997). However, only the relationship between handwriting and the gender of the writer has been validated by experiments (Beech and Mackintosh, 2005).

This study was built on the recent advances of the combination of image processing and machine learning techniques. Some new computer science knowledge and information technology were used for analyzing the gender from off-line scanned handwritten images (Bandi and Srihari, 2005). An information system for processing handwriting can partially replace manual analysis. Many systems have been developed for automatic analysis of handwriting. The main applications include writer identification, word spotting, handwriting recognition, and signature verification. Automatic identification of gender from handwriting is still a challenge with only a few contributions (Ibrahim et al., 2014).

Ref. (Liwicki et al., 2011) proposed a method to predict gender from on-line and off-line handwritings by using the Gaussian mixture models as a classifier. The results of classification showed that on-line features produce better performance than off-line features. Shape description method (Sokic et al., 2012) which contains the tangent angle function, curvature function, and Fourier descriptors to analyze the differences between male and female writers. The typical features of male and female handwritings infer a difference in the shape descriptor. Furthermore, the different features calculated from the same word for gender discrimination were presented and addressed Siddiqi

**Corresponding author: Tel.: +0-8620-8411-1161; fax: +0-8620-8411-1188;
e-mail: mcstj@mail.sysu.edu.cn (Jun Tan)

(Siddiqi et al., 2014). Writing samples can reflect such features as slant, curvature, texture, and legibility. The classifiers are artificial neural networks and support vector machines.

For recent advances in gender prediction from handwriting, the 12th International Conference on Document Analysis and Recognition (ICDAR 2013) (Hassaine et al., 2013) had organized a competition for gender recognition from handwriting. The competition attracted 194 participants from both academia and industry. A competition data set containing Arabic and English handwritings, which were produced by 475 writers, was used. This set is a subset of the Qatar University Writer Identification (QUWI) (Al Máadeed et al., 2012) database that was proposed at the International Conference on Frontiers Handwriting Recognition (ICFHR 2012), There were more than 1000 writers approximately distributed among 50% male and 50% female. The competition provided some known features from the handwritten images. The evaluation metric was the percentage of documents that correctly identified the writer gender.

Male and female writers have various writing styles. Related features can be extracted from scanned handwriting images. In this paper, we only consider some known features to be selected optimally for gender prediction in English and Arabic handwriting. Some feature types include direction, curvature, chain code, gradient direction, textural and allographic features (Bulacu and Schomaker, 2007), and wavelet domain local binary patterns (WD-LBP) (Du et al., 2010). For Chinese handwriting, features such as Fourier transform, shape and structure (Tan et al., 2011), Gabor filter (Tan, 1992), normalization-cooperated gradient (Liu, 2007), fragmented edge structure (Wen et al., 2012), and weighted grid micro-structure (Xu et al., 2011) have been explored.

Feature selection is the process of selecting a subset of relevant features from the original large set of features(Ring and Eskofier, 2015). The minimal-Redundancy-Maximal-Relevance (mRMR) (Peng et al., 2005) method is a specific feature selection method based on Mutual Information (MI). However, when the data is complex, mRMR will increase computation complexity in order to improve the correlation between features and the classes. Another problem is inconsistency (Dash and Liu, 2003) as described in the paper. In this paper, the authors show the time complexity of computing with the inconsistency ratio is close to $O(N)$ when the size of the set is N. The rate is also monotonic and has some tolerance to noise. However, it is only available for discrete values. Hence, it needs discretization for continuous features. The process will greatly affect the computation complexity and consume more memory.

To tackle the above drawback of traditional feature selection methods, we propose a new feature selection approach named kernel mutual information (KMI). KMI benefits from kernel function (Wang et al., 2014) and the mutual information (MI) method. Additionally, KMI joins kernel learning, max-relevance, and min-redundancy. In this paper, we use the KMI method to predict the gender of writers from handwritings.

The contributions of this paper are summarized as follows:

1) We applied the KMI feature selection method for gender prediction to handwritings. The results were compared with other feature selection methods. It was observed that the KMI method had a lower computation complexity and a high identification accuracy.

2) The KMI method analyzes various types of features derived from slant, orientation, roundness, and curvatures in handwriting styles. Such features have been proposed by different researchers, they have been proved to be effective by theory and by experimentation. KMI combines these features together to gain the best performance while reducing the dimension.

3) Experiments were performed on two completely different databases involving handwriting word samples in various languages (Arabic, English, Chinese). Analyzing identification performance includes both text-independent and text-dependent methods, and script-independent and script-dependent modes.

The remainder of this paper is organized as follows: Section 2 introduces some handwriting feature extraction methods and previous related work. We also present the preprocessing study, i.e., handwriting segmentation and optical character recognition(OCR). Section 3 presents the KMI method and its feature selection algorithm, which consists of schemes to select optimal squeezed features. We also discuss the implementation issue to choose among several kernels and different classifiers. The results of experiments for gender prediction on various types of data sets are described in Section 4. The data sets include Arabic, English, and Chinese handwriting data sets. In the last section, the conclusions are discussed and the potential direction of research in future research are concluded.

2. Feature Extraction

Our goal is to discriminate the different styles of male and female writings, the flowchart of our approach is shown in Fig. 1. We focused on all types of features in handwritings which can be categorized into geometrical and transformed features. The geometrical features consider the geometrical information of handwritten images, such as the slant, orientation, roundness, curvatures, etc. They are extracted directly from images. The feature value reflects the pattern of a handwriting style. The transformed features are the result of mathematical transformation on original handwritten images, i.e., Fourier feature, Gabor feature, and WD-LBP feature. Hence, the above features are addressed in this section. Moreover, to extract the special features from the handwritten text, we used the related segmentation method to divide the connect components (CC). The goal is to confirm the gender label from handwritings quickly. The optical character recognition (OCR) technique was adopted and described in this section.

2.1. Geometrical Features

Ordinarily, graphologists use geometrical features to measure the characteristics of handwriting. The general metrics give an overall impression of a handwriting style. Thus, graphologists categorize the handwriting in terms of clarity, fine balance, originality, slant, regularity, and continuity. The personality traits are usually associated with stroke quality (Ahmed

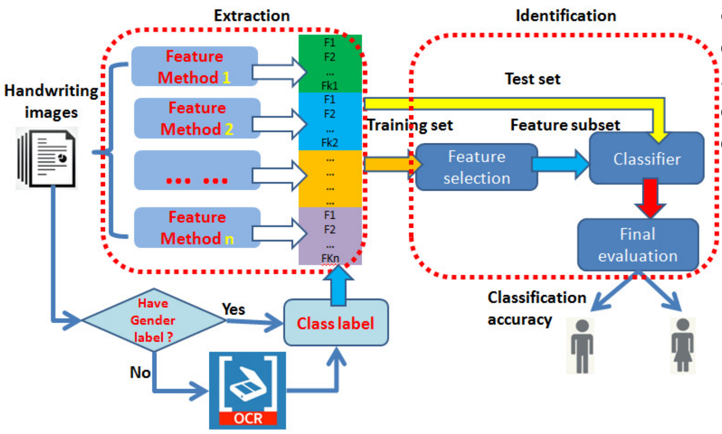


Fig. 1. Flowchart of the approach for gender identification.

and Mathkour, 2008). Related technology has been developed as shown in Fig. 2.

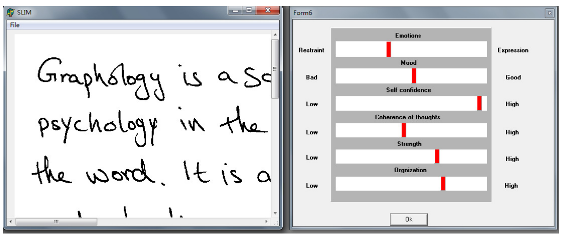


Fig. 2. Analysis of personality traits with handwriting.

The slant and curvature information, extracted from contours, are common features in the analysis of handwriting images. The contour was chosen because of the common belief that the shape of handwritten text, from a handwritten image, can be represented by its contours. Analysis on contours allows us to compensate from the sensitivity of the writing instrument. Contours are expressed by a sequence of chain codes, which have been widely applied to character recognition (Blumenstein et al., 2007) and writer recognition(Siddiqi and Vincent, 2010). Because the work of gender identification also uses handwriting images, we applied the chain code representation for feature extraction. According to the chain code representation, the contour is expressed as a sequence of contour pixels with $\{p_i|1 < i \leq N_{j-1}\}$, where $p_i \in \{0, 1, \dots, 7\}$ and N_j is the length of contour j. Using the chain code, we can calculate the histogram of each code, which ordinarily is seen as a slope density function. The bins of the histogram represent the relative distribution of the eight directions. This is similar to textural and allographic features (Bulacu and Schomaker, 2007). One drawback of using chain code is that the direction between forward and backward strokes are not discriminated.

To measure curvature at the pixel of contour, chain code features calculate the histogram of chain code pairs. We initialized a matrix $A_{8\times8}$ with each element initialized to 0. Once we find a pair $(i, j), 0 \leq i \leq 7, 0 \leq j \leq 7$, we increment the related bin of A. The distribution is regarded as a metric of the curvature of the vector of strokes. The chain code features represent orientation and curvature effectively. However, the feature values are sensitive to noise distortion in writing samples. The chain code features are illustrated in Fig. 3, where C_i is the contour edge between pixel P_{i-1} and P_i, θ_i is an angle of curvature.

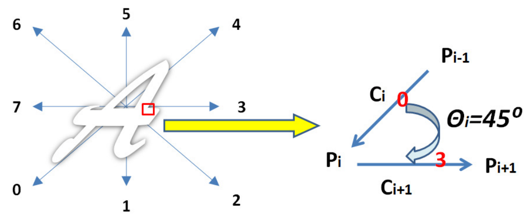


Fig. 3. Chain code representation and pair(0,3) represents a 45° at pixel p_i.

Polygon-base features are used to estimate the contours of handwriting samples by using the sequential polygonization algorithm (Abbasi et al., 2013). The angle interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ is divided into 8 bins. The slope lines, used for approximating the contours, are assigned to their bins respectively. The angle α_i evaluates curvature between each pair of connected segment vectors V_i, V_{i+1}, which is defined as

$$\alpha_i = \pi - \arccos \frac{V_i \cdot V_{i+1}}{|V_i||V_{i+1}|} \tag{1}$$

an example of polygonized contours of handwriting is shown in Fig.4,

Fig. 4. Polygonization algorithm, (a)original handwriting images, and (b) polygonized contours.

The Grid Micro-Structure Feature (GMSF) (Xu et al., 2011) is a text-independent feature which can work in several languages. In addition, the GMSF is extracted from handwriting contours. Suppose the contour image is covered by a $(2N + 1) \times (2N + 1)$ mask and N is the size of the grid. In Fig5, each square in the grid is labeled as i^m, where m denotes the biggest distance from the square to the center, $i \in \{0, \dots, 8m-1\}$ is the square index. The pixel pair$< i^m, j^l >$ represent two contour pixels in a pair of squares. We use $h(i^m, j^l)$ to compute the total number of edge pixels in each square pair of the grid when $< i^m, j^l >$ satisfies the following criteria in Eq.(2):

$$\begin{cases} 1 \leq m = l \leq N, i < j \\ i^m, j^l \in \mathbf{E} \end{cases} \tag{2}$$

where $\mathbf{E}$ is a set of the edge pixels. The positions of the pixel pairs are at the same distance away from the center, the 2 other similar criteria includes $m = l - 1, m = l - 2$.

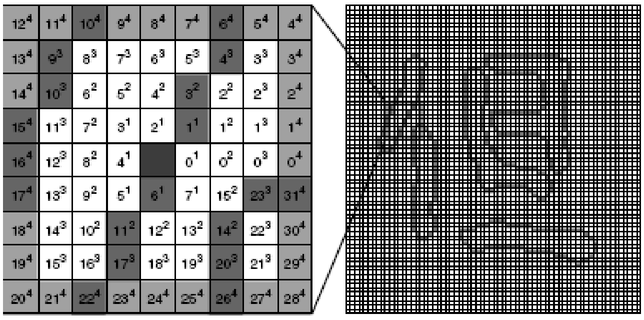


Fig. 5. GMSF features are derived from the pixel pairs.

All of the above conditions are for the three sets and for the pixel pairs $< i^m, j^l >$ in the sets. Hence, $h(i^m, j^l) = h(i^m, j^l) + 1$, and the sum of all occurrence numbers is $H = \sum_{i,j,m,l} h(i^m, j^l)$. The probability density distribution p of the pixel pair $< i^m, j^l >$ is defined as follows:

$$p(i^m, j^l) = \frac{h(i^m, j^l)}{H} \tag{3}$$

where $p(i^m, j^l)$ is named the Grid Microstructure Feature (GMSF).

2.2. *Transformed Feature*

The mathematical transform could be applied to analyse the handwritten images under different scales. Sub-images of different scales reflect the information of the handwritings with various widths, blocks with various size and structure. To get transformed features for gender recognition, we extracted some transform features from handwritten images of different scales.

The Fourier transformation is a traditional method for transforming data from a spatial plane to a frequency plane, which had been applied to handwritten feature extraction in(Chen et al., 2008)(Abandah et al., 2014). For an $N \times N$ given handwriting image $f(x, y)$, the 2D Fast Fourier Transform (FFT) is defined as:

$$F(u, v) = \frac{1}{N} \sum_{x=0}^{N-1} exp[\frac{-j2\pi ux}{N}] \sum_{y=0}^{N-1} f(x, y) exp[\frac{-j2\pi vy}{N}] \tag{4}$$

where $j = \sqrt{-1}$, u, v are the frequency variables. $u, v = 0, 1, \ldots, N-1$, and x, y are the spatial variables. $F(u, v)$ is divided into M parts, and the mean of the energy distribution for each part was computed. We obtained M FFT features as follows:

$$\overrightarrow{F} = [F_1, F_2, \ldots, F_M]^T \tag{5}$$

where $F_i = \sum |F(u, v)_i| / M$, i is an index from part of the FFT energy $|F(u, v)|$ distribution.

The Gabor-filter-based multichannel method (Tan, 1992) has been proven to be the most effective in texture feature extraction for texture classification. Generally in 2D Gabor filers, h_e, h_o denote the even and odd symmetrical filter functions. We define the Gabor filters with the unity aspect ratio as follows:

$$\begin{aligned} h_e(x, y) &= g(x, y) \cdot \cos[2\pi\omega_0(x\cos\theta + y\sin\theta)] \\ h_o(x, y) &= g(x, y) \cdot \sin[2\pi\omega_0(x\cos\theta + y\sin\theta)] \end{aligned} \tag{6}$$

where ω_0 is the central frequency and θ is the orientation. For a given image $f(x, y)$, $g(x, y)$ is one of the output functions of the four input cells which is denoted by

$$g(x, y) = \frac{1}{2\pi\sigma^2} exp[-\frac{x^2 + y^2}{2\sigma^2}] \tag{7}$$

where σ is a spatial constant. $F(u, v)$ is a Fourier transform of $f(x, y)$. Therefore, the Gabor filter is represented in frequency plane as:

$$\begin{aligned} H_e(u, v) &= \frac{[H_1(u, v) + H_2(u, v)]}{2} \\ H_o(u, v) &= \frac{[H_1(u, v) - H_2(u, v)]}{2j} \end{aligned} \tag{8}$$

where $j = \sqrt{-1}$, and the definition of $H_1(u, v), H_2(u, v)$ is given by:

$$\begin{aligned} H_1(u, v) &= exp\{-2\pi^2\sigma^2[(u - \omega_0\cos\theta)^2 + (v - \omega_0\sin\theta)^2]\} \\ H_2(u, v) &= exp\{-2\pi^2\sigma^2[(u + \omega_0\cos\theta)^2 + (v + \omega_0\sin\theta)^2]\} \end{aligned} \tag{9}$$

Based on the above result, for a given $N \times N$ image, we chose six frequencies, ω_0 with $\omega_0 \leq N/2$, and four orientations, i.e, $N = 128$. Hence, $\omega_0 \in \{2, 4, 8, 16, 32, 64\}$ and the four orientations are $\theta \in \{0^o, 45^o, 90^o, 135^o\}$. We obtained 24 central frequency locations defining 24 visual cortical channels. By considering the mean values and the standard deviations of each channel output, we extracted 48 Gabor features from the given texture sample.

2.3. *Handwriting Segmentation*

The above features are the text-independent features, but many features (Tan et al., 2012) for the forensic document examination are text-dependent. Each handwritten document was scanned into a digital picture. Line images and word images were segmented from the document images. Hence, text segmentation was an indispensable step in handwriting recognition - especially for the strong overlapping and touching characters. The separation boundaries between characters are nonlinear, which usually leads to failure for segmentation.

Ref. (Liu et al., 2004) presented a segmentation method for handwritten numbers and texts. Some application results of the methods are shown in Fig.6.

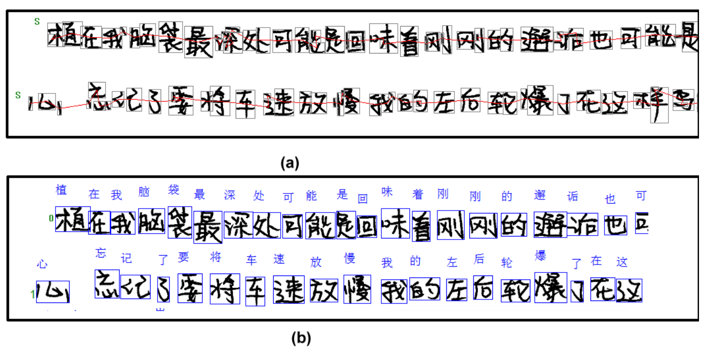


Fig. 6. Examples of handwritten document segmentation:(a)text lines, and (b)Chinese characters.

In their research, all the strokes labeled $\{S_1, S_2, \ldots, S_n\}$ from an image were extracted. n is the total number of strokes. The gravity $< G_x^k, G_y^k >$ of kth stroke S_k is defined as follows:

$$G_x^k = (\sum_{i=1}^{M} i \times P_x(i))/(\sum_{i=1}^{M} P_x(i)) \tag{10}$$

$$G_y^k = (\sum_{i=1}^{N} i \times P_y(i))/(\sum_{i=1}^{N} P_y(i)) \tag{11}$$

where M, N are respectively the width and height of the stroke S_k, and $P_x(i), P_y(i)$ is the pixel in the ith vertical and jth horizontal line. Using spectral clustering, we transformed the Euclidean distance of stroke to a similarity matrix $W = [w_{ij}]^{n\times n}$

$$w_{ij} = \begin{cases} exp((-\|S_i - S_j\|)/\sigma_c) & if\|S_i - S_j\| < r \\ 0 & otherwise \end{cases} \tag{12}$$

where σ_c is the variance of all the distances and r is the mean of all distances. The clustering algorithm was implemented as follows:

1. A diagonal matrix D is calculated by the similarity matrix $W = [w_{ij}]$, $D_{ii} = \sum_{j=1}^{n} w_{ij}$.
2. Gain the matrix $L = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}$.
3. Search the k eigenvectors $\{x_1, x_2, \ldots, x_k\}$ of L, k is the number of clusters, matrix $X = [x_1, x_2, \ldots, x_k]$.
4. Take each row of X as an item and cluster the items into k clusters by K-means.
5. Obtain the cluster label of each stroke according to the cluster label.

Therefore, we obtained the cluster label for the n strokes. They were combined to form characters according to their labels and position.

2.4. Optical Character Recognition

The ICDAR 2013 hassaine competition data set has $475\times 4 = 1900$ handwritten documents. Registration-form documents data set(Tan et al., 2015) has 11118 handwritten documents. Our goal was to identify the gender of the handwriting samples. We needed to obtain the gender labels of the large handwritten documents, which was a huge difficult task in order to transform the handwritten characters into the digital label manually. Hence, optical character recognition (OCR) was used to recognize the gender of the handwritten words. Here, we only need to recognize handwritten Chinese characters of males and females for gender recognition.

In this paper, some handwritten characters of gender from the registration-form documents set are shown in Fig.7. We can see that Fig.7(a) views the characters in the first three rows as males and in the last three rows as females.

Similar to other pattern recognition (Yin and Liu, 2009) methods, the OCR process involves reading an image, converting it to gray scale and black and white images, and resizing the image to that of a template in a reference dictionary, which is shown in Fig.7(b). Here we set a template size of [42, 42] which means that the width and the height of the image are both 42 pixels. We prepare for recognizing a character image matrix A. The recognition scheme for basic characters is based on shape identification. An efficient template matching approach was employed to recognize individual characters of each group. Given the template matrix $T_i, i = 1, \ldots, N$, N is the total number of templates in the dictionary ($N = 2$ in this paper). We computed the correlation between A and Ti and searched the maximum correlation by

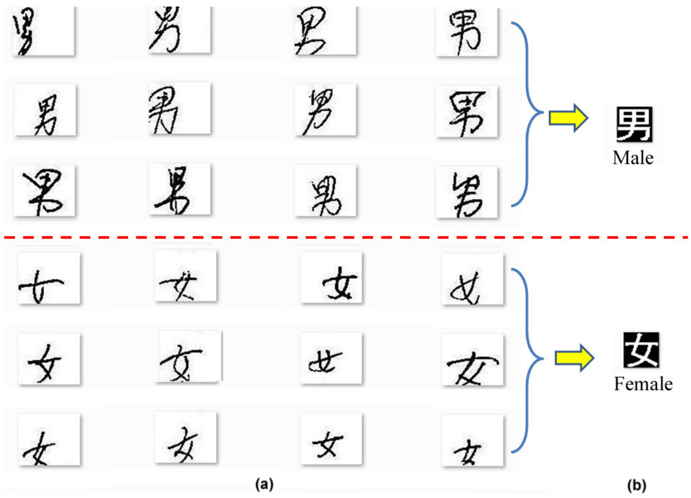


Fig. 7. Chinese handwritten characters express gender,(a)original handwritten characters, and (b) character in reference dictionary .

$$\max_{i\in\{1,\ldots,N\}} \frac{cov(A, T_i)}{\sqrt{cov(A, A)cov(T_i, T_i)}} \tag{13}$$

where $cov(A, T_i)$ is the covariance matrix between A and T_i. According to selecting conditions, and by using the features demonstrated above, we considered the characters in both groups they may be correctly classified.

3. Kernel Mutual Information to Select Feature

The Mutual Information method is an important way to learn mapping of a large number of input features to an output class label. With this view, we integrated the kernel into mutual information. In this section, we review the related methods and present a certain criterion of KMI for feature selection. We performed theoretical analysis of KMI and explain why it is a suitable method.

3.1. Description of Feature Selection Problem

Suppose $\mathcal{X}(\subset \mathbb{R}^{m\times n})$ is a set of input feature vectors. The vector of the output class label is $\mathcal{Y}(\subset \mathbb{R}^m)$. There are m independent distributed paired samples for the u_i class. It is described as follows:

$$\{(u_i, y_i)|u_i \in \mathcal{X}, y_i \in \mathcal{Y}, i = 1, \ldots, m\} \tag{14}$$

here m is number of samples, and it is also the dimension size of the input feature vectors. n is the number of features in set $\mathcal{X}$.

The relationship between m and n will affect the classification performance. x_i is the feature vector with m dimensions. The original data are denoted by:

$$\mathcal{X} = [u_1, \ldots, u_m]^\top = [x_1, \ldots, x_n] \in \mathbb{R}^{m\times n}, \quad (15)$$

$$\mathcal{Y} = [y_1, \ldots, y_m]^\top \in \mathbb{R}^m, \quad (16)$$

where $\top$ means the transpose of a vector. The joint distribution function of density $p_{x,y}(x, y)$ can be derived from here.

In the pattern recognition or machine learning fields, the most important work is to search for a predictor function $f(\cdot)$, which maps the feature vector x to output class y. $f(\cdot)$ is drawn from the training set, including the sample pairs(x, y), by using a machine learning method. If we assume that $f(\cdot)$ is a member function of the predictor categorized by different parameters ω, then we specify a special member function as $f_\omega(\cdot)$. The goal is to use ω minimize the objective function as follows:

$$\min \Gamma(\omega) = \min \int \mathcal{L}(f_\omega(x), y) p(x, y) dxdy \quad (17)$$

where $\mathcal{L}(f_\omega(x), y)$ is a loss function when returning $f_\omega(x)$. In fact, the value y, $p(x, y)$ is a joint density function of x and y. The maximum loss function $\mathcal{L}$ is defined as the misclassification rate for classification problems. The ordinary criterion is in following form:

$$\mathcal{L}(f_\omega(x), y) = \frac{1}{2}\|y - f_\omega(x)\alpha\|_2^2 + \lambda\|\alpha\|_1, \quad (18)$$

where $\alpha = [\alpha_1, \ldots, \alpha_n]$ is a coefficient vector for regression to the features $\lambda > 0$, which is a regularization parameter. $\|\cdot\|_1, \|\cdot\|_2$ are ℓ_1- and ℓ_2- norms.

3.2. Mutual Information for Max-Relevance and Min-Redundancy

The goal of mutual information is to look for one variable and test its value. As a result, the best separation among the classes could be found. To realize the goal, we reduce the class label entropy according to the class distribution function of the training data by the average entropies of the new partitions. Their mutual information $I(x, y)$ of x and y is defined with their individual or joint density probability function $p(x), p(y), p(x, y)$:

$$I(x, y) = \int\int p(x, y) log \frac{p(x, y)}{p(x)p(y)} dxdy \quad (19)$$

The purpose of using MI in feature selection focuses on two issues. The first one is to remove irrelevant variables, named Max-Relevance. The second issue is to remove redundant variables as much as possible, named Min-Redundant. Max-Relevance means select features, via the maximal relevance criterion, between the input feature subset $\Phi, (x_i \in \Phi)$ and the output class y it is denoted by

$$I(\Phi, y) = \frac{1}{|\Phi|} \sum_{x_i \in \Phi} I(x_i, y). \quad (20)$$

It is clear that relevancy requires evaluating a joint metric of redundancy because Max-Relevance usually has a rich redundancy in the feature selection process. Min-Redundant views that two features highly depend on each other. If one of them was removed, the respective classification performance would not decrease, so the following minimal redundancy criterion can be added to choose suitable features in a candidate feature set Φ:

$$R(\Phi) = \frac{1}{|\Phi|^2} \sum_{x_i, x_j \in \Phi} I(x_i, x_j). \quad (21)$$

Combining the above Eq. (20)(21), the criterion is named "minimal-Redundancy-Maximal-Relevance"(mRMR)(Ding and Peng, 2005). Therefore, the optimization problem from Eq.(19) transforms to a new criterion by MI. The following form optimizes $I(\Phi, y)$ and $R(\Phi)$ together:

$$\begin{aligned} \min \mathcal{L}(f_\omega(x), y) &= \min(-I(\Phi, y) + R(\Phi)) \\ &= \max(I(\Phi, y) - R(\Phi)) \\ &= \max[\sum_{x_i \in \Phi} I(x_i, y) - \frac{1}{|\Phi|} \sum_{x_i, x_j \in \Phi} I(x_i, x_j)] \end{aligned} \quad (22)$$

In fact, the MI method is based on a sufficient statistic distribution. Only a low dimensional space may exist for generation of sufficient statistics. There is an absence of sufficient statistics in high dimensions. Therefore, evaluating the MI could be difficult in order to solve the optimization problem in the general multi-class high dimension feature space.

3.3. Integrate Kernel Function with MI

The Kernel function maps the original data to a high dimensional feature space. With the Kernel transform, the optimization problem of the redundancy and relevance of features may be solved easily. Widely used types of kernel functions, including the polynomial kernel, Gaussian kernel (or called radial basis function (RBF) kernel), exponential kernel, sigmoid kernel, and delta kernel:

Polynomial kernel:

$$k(x, y) = (x^\top y + c)^d, \quad (23)$$

Gaussian kernel:

$$k(x, y) = exp(-\frac{\|x - y\|^2}{2\sigma^2}), \quad (24)$$

Exponential kernel:

$$k(x, y) = exp(-\frac{\|x - y\|}{2\sigma^2}), \quad (25)$$

Sigmoid kernel:

$$k(x, y) = tanh(cx^\top y + d), \quad (26)$$

where c,d and σ are parameters of the kernels. Their values are obtained from practical problems. The value of *sigma* is usually set to the number of samples.

We consider the feature vector x, a mapping function $\phi(\cdot) : x \longrightarrow \phi(x)$, and a transform output vector y using the mapping $\phi(\cdot) : y \longrightarrow \phi(y)$. $\phi(\cdot)$ is a nonlinear mapping function and a kernel function can be denoted by the mapping function: $k(x, y) = \phi(x)^\top \phi(y)$. Therefore, the optimization equation for feature selection from Eq.(18), using the kernel method, is given by

$$\min \mathcal{L}(f_\omega(x), y) = \min\{\|\phi(y) - \alpha\phi(x)\|_F^2 + \lambda\|\alpha\|_1\} \tag{27}$$

where $\|\cdot\|_F$ is Frobenius norms. Then, we may see that:

$$\begin{aligned}\mathcal{L}(f_\omega(x), y) &= \|\phi(y) - \alpha\phi(x)\|_F^2 + \lambda\|\alpha\|_1\\ &= Tr(\phi(y) - \alpha\phi(x))^\top(\phi(y) - \alpha\phi(x)) + \lambda\|\alpha\|_1\\ &= Tr(\phi(y)^\top\phi(y)) - 2Tr(\alpha^\top\phi(y)^\top\phi(x))\\ &\quad + Tr(\alpha^\top\phi(x)^\top\phi(x)\alpha) + \lambda\|\alpha\|_1\end{aligned} \tag{28}$$

To optimize the above equation, we also assumed that the loss function $\mathcal{L}$ for the given input data was differentiable with respect to the coefficient vector α, i.e. we may compute $\partial\mathcal{L}(x_i, y)/\partial\alpha$ for the willful feature vector x_i followed by the willful coefficient α_i. We have:

$$\forall\alpha, \frac{\partial\mathcal{L}}{\partial\alpha} = 0, \tag{29}$$

Hence, the gradient of the criterion with respect to the arbitrary coefficient α is:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\alpha} &= \frac{\partial[Tr(\phi(y)^\top\phi(y))]}{\partial\alpha} - \frac{\partial[2Tr(\alpha^\top\phi(y)^\top\phi(x))]}{\partial\alpha}\\ &+ \frac{\partial[Tr(\alpha^\top\phi(x)^\top\phi(x)\alpha)]}{\partial\alpha} + \frac{\partial[\lambda\|\alpha\|_1]}{\partial\alpha}\\ &= -\phi(x)^\top\phi(y) + |\alpha|^\top\phi(x)^\top\phi(x) + \lambda\\ &= -k(x, y) + |\alpha|^\top k(x, x) + \lambda\end{aligned} \tag{30}$$

where we considered the case $\lambda = 0$, compared to Eq.(22). The criterion above, defined with the kernel, is similar to the mutual information. The kernel method was used to test the dependence between two random variables. Hence, we instead define MI $I(x, y)$ with kernel function $I_k(x, y)$, called the Kernel Mutual Information (KMI) function $I_k(\cdot)$. We obtain the new criterion:

$$\begin{aligned}\min \mathcal{L}(f_\omega(x), y) &= \max(I_k(\Phi, y) - R_k(\Phi))\\ &= \max[\sum_{x_i\in\Phi} I_k(x_i, y) - \frac{1}{|\Phi|}\sum_{x_i\in\Phi; x_j\notin\Phi} I_k(x_i, x_j)]\end{aligned} \tag{31}$$

where R_k is a function redefined by the kernel from Eq.(21) and Φ is a feature set which has i features selected. The purpose of the method is to find the best optimal $(i+1)$th feature from set $\{X - \Phi\}$. The respective optimization algorithm to search for the $(i+1)$th feature can be denoted by:

$$\max_{x_i\in\Phi}[I_k(x_k, y) - \frac{1}{i}\sum_{x_j\in X-\Phi; x_i\in\Phi} I_k(x_j, x_i)] \tag{32}$$

where the parameter $|\Phi|$ has been replaced by i which is the number of selected features. This criterion forms the basis for later feature selection algorithms .

4. Experiment

In this section, we used our proposed KMI method to select an optimal subset of features in order to identify the writer gender in two handwritten data sets. The data sets have multi-features and samples and the total size of the features in the two data sets is large. The ranges of the feature values vary. Thus various data sets increase the challenges to gender identification. It would require a quality feature selection method to achieve a higher accuracy and be robust in complexity experimental conditions. We compared our proposed KMI methods with existing methods for gender prediction on above data sets. The comparison results of performance are discussed and analyzed.

4.1. Data Sets

Two real handwritten data sets were used in our experiments for gender prediction which include the ICDAR 2013 sets(Hassaine et al., 2013) and our registration-form document data sets(Tan et al., 2015).

ICDAR 2013 (Hassaine et al., 2013) organized a competition for gender identification from handwriting through Kaggle. The data set is publicly available and has more than 475 writers with a distribution of about 50% male and 50% female. Each writer wrote four handwritten documents - two pages containing Arabic handwritten text and another two pages containing English handwritten text. The competition provided geometric features extracted from all the handwritten images. All images were digitized by an Epson GT-S80 scanner at a 600 DPI resolution. The format of the digital images is JPEG. The training set has 282 writers whose genders are provided. The gender of the remaining 193 writers is required to predict.

Our registration-form documents (RDF) set has 11 118 handwritten documents which consist of 9 256 gray images and 1 862 color images. In this paper, we consider each type of field in the registration forms except for the sheet letters, which were machine-printed. All other words were handwritten. All the writers were required to write his or her name and personal information on the registration form. Because the form consisted of personal information, the data set is not open to the public. The text of the form contains Chinese characters and Arabic numerals. Personal information (address, workstation, gender, birthday, telephone number, post code, person ID, job) was only written once. Handwritten signature, Arabic numerals, and corresponding Chinese characters were used frequently. The writers were requested to write the content five times in the forms. In our experiments, we selected 75% form images for training, and the remaining images were used as the test set.

The descriptions of the data sets are summarized below (shown in Table1):

The data stored in the ICDAR 2013 database consists of feature values and original document images. The RDF data set only has original images. Some sample images of the above data set are shown in Fig.8. Besides the feature values, the ICDAR 2013 training data set contains the class label. The value of the class label could be 0 or 1. A value of '1' indicated male and '0' for female.

Many feature types were considered for gender identification in this paper. We used the KMI feature selection algorithm for the multi-feature problem. Some selected features were geometry-based features and others were transformation features. We list the various feature types in Table 2.

Table 1. Summary of real handwritten data sets.

Data Set Name	Language	# of Documents	# of Writers	# of Features	Is Public
ICDAR 2013(Hassaine et al., 2013)	Arabic+English	1900	475	7066	Yes
RDF (Tan et al., 2015)	Chinese	11118	11118	9517	No

Table 2. Summary of feature types.

Feature Type Name	#of Features	Transformed or Geometrical	Text-Independent	Data Set
TortuosityHist (Abbasi et al., 2013)	10	Geometrical	Yes	ICDAR 2013
Direction (Blumenstein et al., 2007)	40	Geometrical	Yes	ICDAR 2013
Curvature (Siddiqi and Vincent, 2010)	900	Geometrical	Yes	ICDAR 2013
Chain Code (Siddiqi and Vincent, 2009)	5020	Geometrical	Yes	ICDAR 2013
Gradient Direction (Liu, 2007)	1096	Transformed	Yes	ICDAR 2013
Fourier (Chen et al., 2008)	1600	Transformed	Yes	RDF
Gabor (Tan, 1992)	36	Transformed	Yes	RDF
GMSF (Li and Ding, 2009)	6561	Geometrical	Yes	RDF
Allographic (Bulacu and Schomaker, 2007)	1296	Geometrical	Yes	RDF
Shape+Structure (Tan et al., 2011)	24	Geometrical	No	RDF

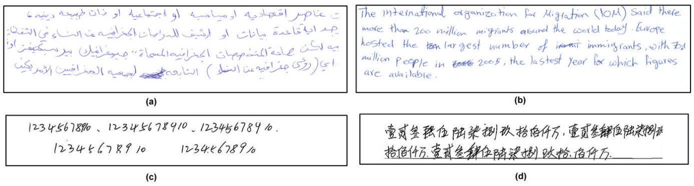


Fig. 8. Some original images in ICDAR 2013 and RDF, (a)Arabic text, (b)English text, and (c) Arabic numbers, (d)Chinese text.

4.2. Comparison of Methods

The goal of this paper is to employ feature selection to optimize the subset for gender recognition. According to the features that have been introduced above, we needed to prove that the KMI feature selection method would improve the accuracy more than using the isolated features. In this experiment, we chose the kernel and classifier in order to finish the task effectively. Furthermore, we compared our proposed KMI methods with the existing methods using the above data sets. Comparison results of the performances are addressed.

As commonly accepted, the kernel metric is the bridge that connects the original data with the learning methods. Kernel methods play a key role for KMI and they are convenient. Theoretically, kernels should be adapted to the requirements of real applications for all types of cases. The ideal kernel function ϕ is denoted as $k(x, y) = \phi(x)\phi(y)$, but we do not know the form of the function $\phi(\cdot)$. Learning is required in most real cases. Therefore, we begin with a simple kernel and combine it to more complex kernels. Some kernels come from different families and are chosen for use in many applications. We analyzed the properties of various kernels from different kernel families. In this experiment, four popular kernels including linear, polynomial, RBF, and Tanh were considered.

Fig. 9 shows the mean classification accuracy of gender prediction using KMI with various kernels. The Tanh kernel achieved the highest performance in two data sets. The highest accuracy was obtained on the RDF data set. We had an average accuracy of 67.2%. The difference between Tanh and the other kernels was small. The linear kernel had the lowest average accuracy at 61.7%. The accuracies of all kernels were similar. Performance of the Tanh kernel was 67.2% and 66.3% in RDF and ICDAR 2013 sets respectively. Achievements of the other kernels were less than Tanh. As a result, we decided to choose the Tanh kernel as our KMI kernel according to the requirements of gender recognition from handwriting.

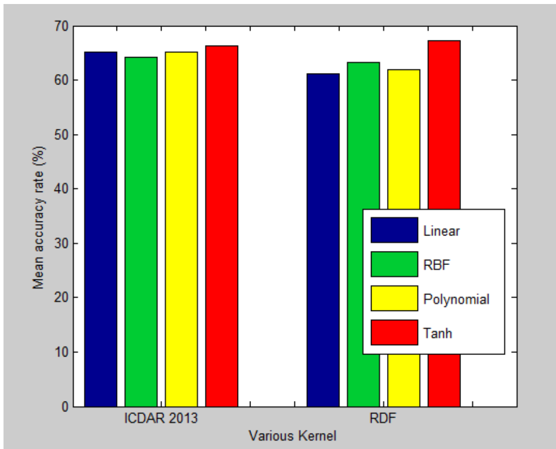


Fig. 9. Mean performance of various kernels.

In the KMI framework, the classifier plays a vital role. Hence, we wanted to choose a classifier that would perform feature selection to provide a good performance. Three popular classifiers were considered which consisted of Support Vector Machine (SVM), Naive Bayes (NB), and k-Nearest Neighbor (kNN).

Fig. 10 shows the mean classification accuracies for NB, SVM, and kNN. As can be seen, the accuracy for the SVM achieved 67.2%, which was higher than both NB and kNN. The mean accuracy of kNN was close to that of the SVM. The highest accuracy for the kNN was 63.8% on the RDF data set and the lowest was 59% on the ICDAR 2013 data set. The corresponding accuracy for the SVM was 66.3%.

Compared with the kNN and SVM, NB did not perform as well in these experiments. The NB classifier requires calculus density and probability, and the values of features to be positive. Thus, occasionally, the NB classifier could not be implemented normally. The mean classification accuracy of the NB was the lowest among the three classifiers. The corresponding accuracies of the NB were respectively 49.0% and 54.0% in the two data sets. Therefore, in the following experiment, SVM was chosen as the default classifier.

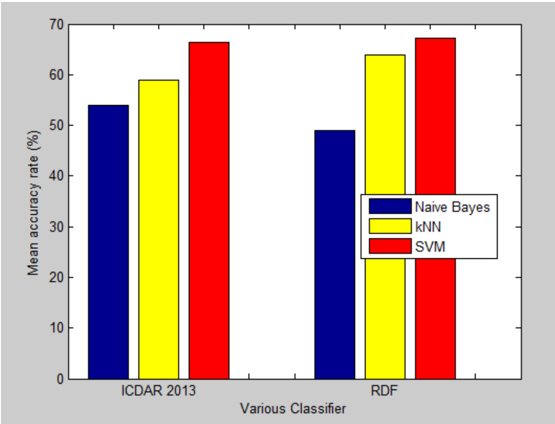


Fig. 10. Mean performance of various classifiers.

Two real world handwritten data sets were used in our experiments - ICDAR 2013 and RDF. Different feature types were used to evaluate the performance of gender identification. Comparisons among the results from the two data sets are summarized below (shown in Tables (3),(4)):

- Table 3 views the average individual classification accuracies for the different feature types in ICDAR 2013. The first set type is TortuosityHist(Abbasi et al., 2013). The size of the features is 10, which is the smallest feature set among ICDAR 2013. Accuracy for this type of feature was 59.2%. The performance of the Direction (Blumenstein et al., 2007) features had a high accuracy at 61.7%. Classification accuracy of Curvature (Siddiqi and Vincent, 2010) was lowest, indicating that the Direction (Blumenstein et al., 2007) features were better than the Curvature (Siddiqi and Vincent, 2010) features. The size of the chain code features was the largest among all the types. The number of features is 5020, but many values are zero. When all types of features are combined, the number of features becomes 7066. Accuracy for the combined features was 65.2%. We used the KMI method for gender identification and we obtained the highest classification accuracy, from the KMI, at 67.3%. When the number of features was 150, we had the lowest accuracy of KMI at 64%. When the number of features was 10, the mean accuracy was 66.3%.

- From Table 4 show that the experimental results obtained from using the RDF set. The size of the Gabor (Tan, 1992) type features was the smallest, at 36. However, the accuracy of Gabor (Tan, 1992) was highest among all types of features, which was 56.7%. The size of GMSF (Li and Ding, 2009) type of features was the largest, at 6561. The lowest accuracy was 35.2% when we only used Fourier features. The accuracy was 61.1% when we combined all the features together. The total number of all features was 9517. The highest accuracy of KMI was 66.7% when the number of selected features was 200 and the lowest accuracy was 38% when the selected feature number was 10. The KMI accuracy has the best performance among the other methods. Results from Table4 show that, for the main data set, the mean classification accuracies of KMI were significantly higher than the individual and total features.

Table 3. Result of gender recognition in ICDAR 2013, ICDAR feature performance shown for comparison.

Feature Type	# of Features	Accuracy
TortuosityHist (Abbasi et al., 2013)	10	59.2%
Direction (Blumenstein et al., 2007)	40	61.7%
Curvature (Siddiqi and Vincent, 2010)	900	56.1%
Chain Code (Siddiqi and Vincent, 2009)	5020	58.5%
Gradient Direction (Liu, 2007)	1096	58.8%
Total features	7066	65.2%
KMI	150	66.3%

Table 4. Result of gender recognition in RDF, all types of features are performed for comparison.

Feature Type Name	#of Features	Accura
Fourier (Chen et al., 2008)	1600	35.2%
Gabor (Tan, 1992)	36	56.7%
GMSF (Li and Ding, 2009)	6561	37.8%
Allographic (Bulacu and Schomaker, 2007)	1296	45.9%
Shape+Structure (Tan et al., 2011)	24	37.8%
Total features	9517	61.1%
KMI	200	66.7%

We selected various numbers of features and employed them in this experiment in order to validate our algorithm. Therefore, we selected the top $10, 30, 50, 100, 150, 200, 400, 600, and 800$ features respectively from the ICDAR 2013 and RDF data sets. We see in Fig. 11, the red curve, representing the accuracy of gender recognition in the ICDAR 2013 data set, and the blue curve that describes the accuracy of gender recognition in the RDF data set. The result shows the lowest accuracy was 38% when the number of features was 10 in RDF. KMI had the highest accuracy at 68% when the top selected number was 200. The accuracy did not change for the number of features of $\geq$ 800. In Fig. 11, the left and right axes show the performances in the various data sets. The left blue axis is for RDF and the right red vertical axis expresses the classification accuracy of KMI in ICDAR, which is high. Results using KMI achieved a 67.3% accuracy rate when the selected number of features was 150. We also selected the top $1000, 2000, 4000, 6000$ features and discovered that when the number of features is greater than 200, the recognition accuracy will stay constant at 65%. With

Table 5. Mean accuracy of gender recognition comparison.

Algorithm	ICDAR 2013	RDF
mRMR(Peng et al., 2005)	63.7%	61.4%
Correlation	62.5%	59.3%
KMI	66.3%	66.7%

Table 6. Mean running time (second/per feature).

Algorithm	ICDAR 2013	RDF
mRMR(Peng et al., 2005)	0.9314	0.6213
Correlation	1.1685	0.8125
KMI	0.7778	0.564

various numbers of selected top features, the experiment can be tested in small-sample or big-sample conditions. All of this study was to promote objectivity when comparing all the methods.

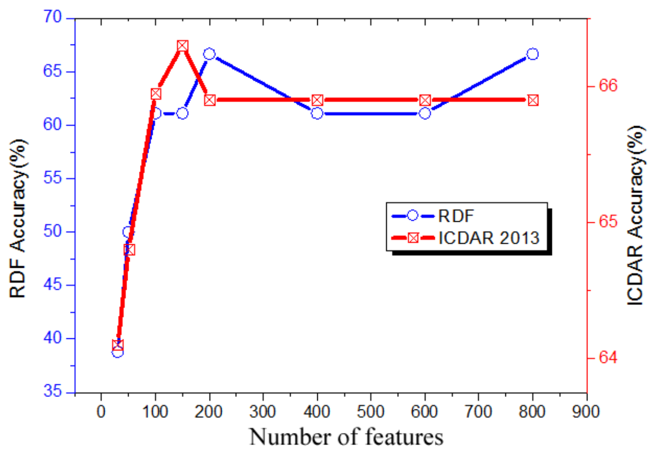


Fig. 11. Gender identification accuracy of KMI in various data sets.

The KMI, mRMR, and Correlation methods are approximations of the mutual information selection scheme, their classification accuracies are shown in Table 5. Thus, we only compared their performances in terms of computation complexity. These comparisons are expressed by the running time for real data. We know that mRMR cannot be executed in complex data sets, therefore, we compared them using the ICDAR 2013 and RDF data sets.

Experiments were conducted on an Intel Core i5-4250 CPU machine with 4 GB memory running windows 7.0. The software was developed and executed with Matlab versions 7.0 and v2012.

We obtained the computation complexity by computing the average time for one feature. The running time was obtained by running all methods and selecting the top m features for each data set. We then recorded the total runtime T and calculated the average running time to be m/T seconds/per feature. Table 6 shows the time costs for the three methods and for each of the data sets. In ICDAR 2013, all methods took longer to finish the task than in ICDAR. KMI took about 0.778 seconds to select one feature in ICDAR 2013. The time for mRMR and Correlation are 0.913 and 1.16 seconds respectively.

We also tested computation complexity in the RDF data set, which was hard work for the mRMR. In ICDAR 2013, the Correlation method took 0.8125 seconds and KMI took 1.833 seconds. Correlation took an average of 0.8125 seconds to find a feature while KMI uses 1.16 seconds in ICDAR 2013. The data set not only had the largest number of features, their feature values were more complex than the other data sets. As a result, the time to run them using ICDAR 2013 was longer than that for the RDF data set.

From the above data analysis, and our experiment in the two data sets, KMI had the lowest average running time compared to mRMR and Correlation. The result demonstrates the effectiveness of the KMI method in computation complexity.

4.3. Analysis of Results

With our proposed KMI methods, we focused on gender prediction based on handwritings from multi-featured data sets. KMI is a feature selection method which consists of a kernel function and classifier. It makes use of the wrapper filter process to select the features. We obtained a high accuracy performance at a low computation cost. The KMI method optimizes the mutual information to yield a candidate feature subset by using a kernel function. Thus, the features in the subset have new metrics based on max-relevance and min-redundancy. KMI is ideal for solving the problem of multi-feature, high dimensional, and complexity of feature values.

Compared to traditional gender recognition algorithms, KMI avoids the conflict of multi-features. The task is very challenging and it is easy to make mistakes when multi-features are mixed together. According to the analysis and experiments shown in the previous section, we used kernel-based methods to compute the mutual information of gender label and feature vectors. The result proved that the difficulty of using multi-features could be overcome. Kernel-based methods can improve the accuracy and reduce the redundancy effectively.

Although, in our experiments, KMI feature selection usually had a higher classification accuracy than the other methods, occasionally, there were some ups and downs in the classification accuracy and redundancy rate. For instance, in Tables 4 and 6, KMI did not have the highest performance when the number of features was minimum. Many conditions can be the reasons for the fluctuations. The first reason was that not enough features had led to bad classification results. Another reason was that cross validation may have yielded the fluctuations in classification.

Traditional feature selection methods consist of incremental search, which includes both forward and backward selections. All these methods cannot confirm the local optimization of the search path since the main difficulty in searching is to cover the entire vector space. Meanwhile, global optimization leads to over-fitting and high computational cost. Hence, KMI obtained a greater classification accuracy and good performance while having a low computation cost.

In our experiments, our goal was to compare the various kernels, classifiers, and other gender classification methods. In the most complex conditions, KMI increased the accuracy of classification and decreased the computational cost. It had the highest classification accuracy in the ICDAR 2013 data set. Although partial achievement on RDF was not satisfactory, it has the overall highest accuracy and it solved the problem the other methods could not perform on this data set. The results of our experiment show that the KMI performs with high accuracy and is an effective method for feature selection to predict gender. The KMI method can be considered as an open, innovative framework, to select features with high effectiveness and to merge all possible new features.

5. Conclusion

In this paper, we used the KMI feature selection method to identify writer gender from handwriting. Under this framework, various types of geometrical and transformed multi-features were mixed together. The features came from different sources. The goal of this paper was to identify gender using multi-features. Compared with other methods, the KMI method improved the accuracy of gender identification and avoided the complexity computation problem. Finally, we reported the results of our experiments using two data sets. More than 7,066 mixed features in ICDAR 2013 data set and 9517 features in RDF were used. The average rate of correct classification was higher than the compared methods. The KMI method integrates the kernel function, classifier, and mutual information. Results of our experiments demonstrated that the KMI candidate feature search, from the remaining feature subset, performed very well at low processing cost.

6. Acknowledgments

This work was supported by the Guangdong Provincial Government of China through the ”Computational Science Innovative Research Team” program and the Guangdong Province Key Laboratory of Computational Science at the Sun Yat-Sen University, the Technology Program of GuangDong (Grant No. 2012B091100334), and the National Science Foundation of China (Grant No. 11471012), and China Scholarship Council(Grant No. 201506385010).

References

Abandah, G.A., Jamour, F.T., Qaralleh, E.A., 2014. Recognizing handwritten arabic words using grapheme segmentation and recurrent neural networks. International Journal on Document Analysis and Recognition (IJDAR) 17, 275–291.

Abbasi, H., Olyaee, M., Ghafari, H.R., 2013. Rectifying reverse polygonization of digital curves for dominant point detection. International Journal of Computer Science Issues(IJCSI) 10, 154–163.

Ahmed, P., Mathkour, H., 2008. On the development of an automated graphology system., in: Proceedings of International Conference on Artificial Intelligence, pp. 897–901.

Al Máadeed, S., Ayouby, W., Hassaine, A., Aljaam, J.M., 2012. Quwi: An arabic and english handwriting dataset for offline writer identification, in: Proceedings of International Conference on Frontiers in Handwriting Recognition (ICFHR), IEEE. pp. 746–751.

Bandi, K., Srihari, S.N., 2005. Writer demographic classification using bagging and boosting, in: Proceedings of 12th International Conference on Graphonomics Society, pp. 133–137.

Beech, J.R., Mackintosh, I.C., 2005. Do differences in sex hormones affect handwriting style? evidence from digit ratio and sex role identity as determinants of the sex of handwriting. Personality and Individual Differences 39, 459–468.

Blumenstein, M., Liu, X.Y., Verma, B., 2007. An investigation of the modified direction feature for cursive character recognition. Pattern Recognition 40, 376–388.

Bulacu, M., Schomaker, L., 2007. Text-independent writer identification and verification using textural and allographic features. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 701–717.

Chen, Q., Yan, Y., Deng, W., Yuan, F., 2008. Handwriting identification based on constructing texture, in: Proceedings of First International Conference on Intelligent Networks and Intelligent Systems(ICINIS), IEEE. pp. 523–526.

Dash, M., Liu, H., 2003. Consistency-based search in feature selection. Artificial Intelligence 151, 155–176.

Ding, C., Peng, H., 2005. Minimum redundancy feature selection from microarray gene expression data. Journal of Bioinformatics and Computational Biology 3, 185–205.

Du, L., You, X., Xu, H., Gao, Z., Tang, Y., 2010. Wavelet domain local binary pattern features for writer identification, in: Proceedings of 20th International Conference on Pattern Recognition (ICPR), IEEE. pp. 3691–3694.

Hassaine, A., Al Maadeed, S., Aljaam, J., Jaoua, A., 2013. Icdar 2013 competition on gender prediction from handwriting, in: Proceedings of 12th International Conference on Document Analysis and Recognition (ICDAR), IEEE. pp. 1417–1421.

Ibrahim, A.S., Youssef, A.E., Abbott, A.L., 2014. Global vs. local features for gender identification using arabic and english handwriting, in: Proceedings of IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), IEEE. pp. 000155–000160.

Li, X., Ding, X., 2009. Writer identification of chinese handwriting using grid microstructure feature, in: Advances in Biometrics. Springer, pp. 1230–1239.

Liu, C.L., 2007. Normalization-cooperated gradient feature extraction for handwritten character recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 1465–1469.

Liu, C.L., Sako, H., Fujisawa, H., 2004. Effects of classifier structures and training regimes on integrated segmentation and recognition of handwritten numeral strings. IEEE Transactions on Pattern Analysis and Machine Intelligence 26, 1395–1407.

Liwicki, M., Schlapbach, A., Bunke, H., 2011. Automatic gender detection using on-line and off-line information. Pattern Analysis and Applications 14, 87–92.

Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 1226–1238.

Ring, M., Eskofier, B.M., 2015. Optimal feature selection for nonlinear data using branch-and-bound in kernel space. Pattern Recognition Letters 68, 56–62.

Siddiqi, I., Djeddi, C., Raza, A., Souici-meslati, L., 2014. Automatic analysis of handwriting for gender classification. Pattern Analysis and Applications , 1–13.

Siddiqi, I., Vincent, N., 2009. A set of chain code based features for writer recognition, in: Proceedings of 10th International Conference on Document Analysis and Recognition(ICDAR), IEEE. pp. 981–985.

Siddiqi, I., Vincent, N., 2010. Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. Pattern Recognition 43, 3853–3865.

Sokic, E., Salihbegovic, A., Ahic-Djokic, M., 2012. Analysis of off-line handwritten text samples of different gender using shape descriptors, in: Proceedings of IX International Symposium on Telecommunications (BIHTEL), IEEE. pp. 1–6.

Tan, J., Lai, J.H., Wang, C.D., Feng, M.S., 2011. A stroke shape and structure based approach for off-line chinese handwriting identification. International Journal of Intelligent Systems and Applications (IJISA) 3, 1.

Tan, J., Lai, J.H., Wang, P., Bi, N., 2015. Multiscale region projection method to discriminate between printed and handwritten text on registration forms. International Journal of Pattern Recognition and Artificial Intelligence 29, 153–185.

Tan, J., Lai, J.H., Zuo, X.X., 2012. The dataset system of economic dispute handwritten (dsedh) based on stroke shape and structure features, in: Proceedings of 21st International Conference on Pattern Recognition (ICPR), IEEE. pp. 661–664.

Tan, T., 1992. Texture feature extraction via visual cortical channel modelling, in: Proceedings of 11th IAPR International Conference on Pattern Recognition Vol. III. Conference C: Image, Speech and Signal Analysis, IEEE. pp. 607–610.

Tett, R.P., Palmer, C.A., 1997. The validity of handwriting elements in relation to self-report personality trait measures. Personality and Individual Differences 22, 11–18.

Wang, J.J.Y., Bensmail, H., Gao, X., 2014. Feature selection and multi-kernel learning for sparse representation on a manifold. Neural Networks 51, 9–16.

Wen, J., Fang, B., Chen, J., Tang, Y., Chen, H., 2012. Fragmented edge structure coding for chinese writer identification. Neurocomputing 86, 45–51.

Xu, L., Ding, X., Peng, L., Li, X., 2011. An improved method based on weighted grid micro-structure feature for text-independent writer recognition, in: Proceedings of International Conference on Document Analysis and Recognition (ICDAR), IEEE. pp. 638–642.

Yin, F., Liu, C.L., 2009. Handwritten chinese text line segmentation by clustering with distance metric learning. Pattern Recognition 42, 3146–3157.